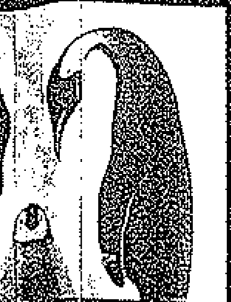
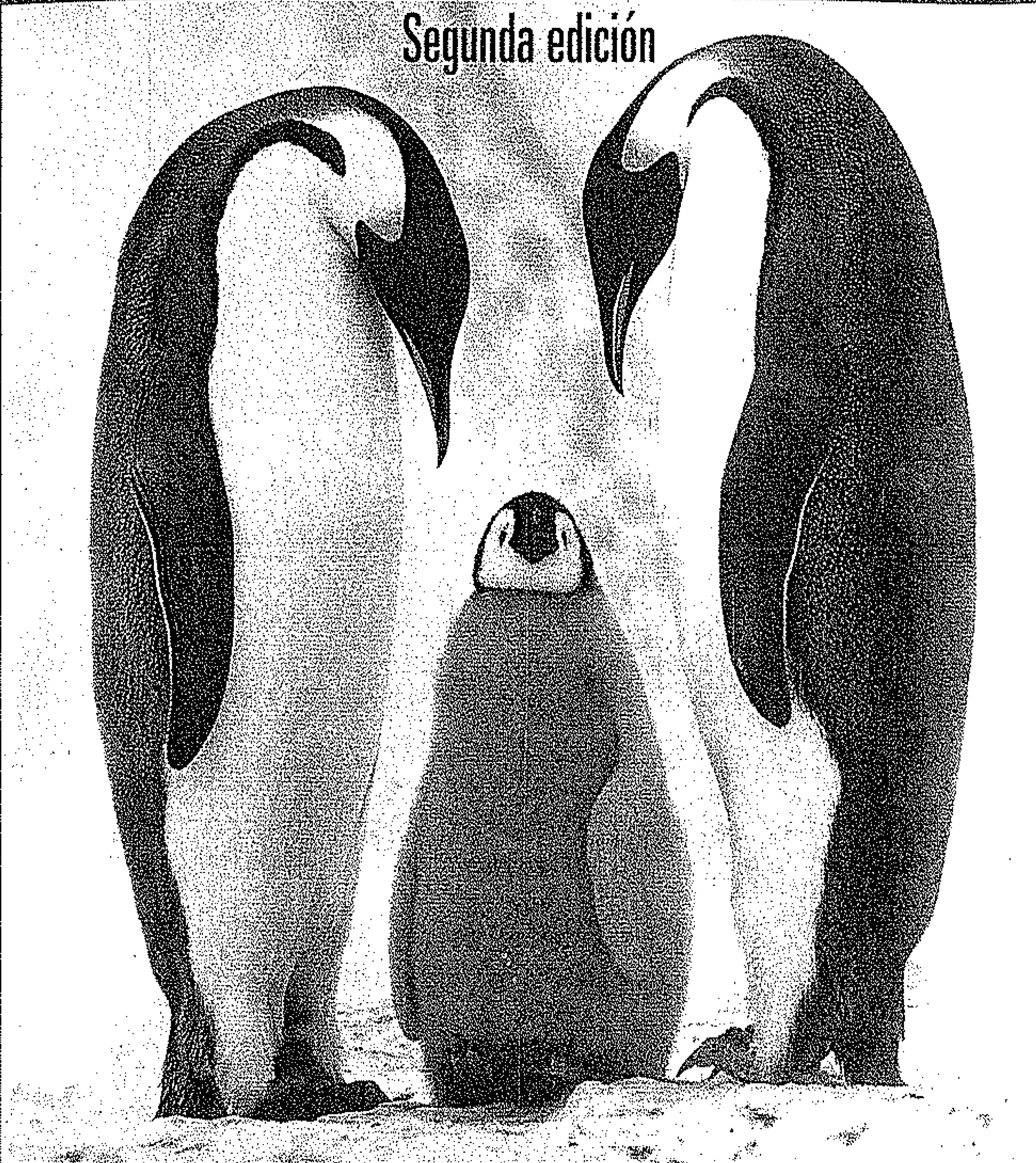


Estadística para psicología

Segunda edición



Prentice Hall

Arthur Aron
Elaine N. Aron

Estadística para psicología
Segunda edición
Arthur Aron
Elaine N. Aron
Prentice Hall

ARTHUR ARON
ELAINE N. ARON

ESTADÍSTICA PARA PSICOLOGÍA

Prentice Hall

Pearson Educación

Argentina • Bolivia • Brasil • Colombia • Costa Rica • Chile • Ecuador • El Salvador • España • Guatemala • Honduras • México • Nicaragua • Panamá • Paraguay • Perú • Puerto Rico • República Dominicana • Uruguay • Venezuela

Amsterdam • Harlow • Londres • Menlo Park • Munich • Nueva Delhi • Nueva Jersey • Nueva York • Ontario • París • Sidney • Singapur • Tokio • Toronto • Zurich

Datos de catalogación bibliográfica

519.5 ARO

Aron, Arthur
Estadística para psicología / Arthur Aron y Elaine Aron. – 1ª ed. – Buenos Aires: Pearson Education, 2001.
736 p.; 19,5x25,5 cm.

Traducción de: Karina Abraham

ISBN: 987-9460-66-9

I. Aron, Elaine II. Título – 1. Estadísticas

Editor: Darío Rubinstein
Gerente de División: Esteban Lo Presti
Armado de Tapa e Interior: Carlos Pérez Villamil / María Rosa Ruggiero
Traducción: Karina Abraham
Corrección: Victoria Aljanati
Producción: Laura G. Lago

Traducido de: Statistics for Psychology, Second Edition by Arthur Aron and Elaine N. Aron, Copyright 1999. Todos los Derechos Reservados. Publicado con el acuerdo del editor original, PRENTICE HALL, INC., Editorial de Pearson Education Company.

ISBN: 0-13-914078-6

Edición en Español publicada por:

Av. Regimiento de Patricios 1959 (C1266AAF), Buenos Aires, Rep. Argentina

ISBN: 987-9460-66-9

Primera Edición: Diciembre 2001
Queda hecho el depósito que dispone la ley 11.723

Impreso en Brasil por RR Donnelley, en el mes de diciembre de 2001.
Rua Epiacaba 90 –Vila Arapuá–
04257.170 São Paulo SP

Tabla de contenidos

2 Media, varianza, desvío estándar y puntuaciones Z 34

3 Correlación 68

7 Pruebas de hipótesis con medias muestrales 202

8 Potencia estadística y tamaño de efecto 232

9 Prueba *t* para medias dependientes 274

10 Prueba *t* para medias independientes 312

11 Introducción al análisis de varianza 344

12 El modelo estructural en el análisis de varianza 376

13 Análisis factorial de varianza 406

14 Pruebas chi-cuadrado 460

15 Estrategias a aplicar cuando las distribuciones poblacionales no son normales: transformación de datos, pruebas de rango y orden y métodos intensivos por computadora 494

16 Integración de contenidos: el modelo lineal general 526

17 Comprensión de los procedimientos estadísticos avanzados que aparecen en publicaciones científicas 558

Prefacio para el profesor

El corazón de la primera edición de este libro fue escrito, durante un verano, en un pequeño departamento de París cerca de Place Saint Ferdinand, y diseñado en los cafés de la zona y durante las caminatas por el Bois de Boulogne. Treinta años de experiencia en la enseñanza, la investigación y la redacción avalan esta obra. Creemos que el libro que logramos es tan diferente de los libros convencionales de estadística como París lo es de Calcuta; es más, incluso consideramos que resultará práctico y estimulante para la sufrida comunidad de profesores de estadística.

El método que da forma al texto se ha ido desarrollando durante tres décadas de enseñanza exitosa, no sólo porque los alumnos continuamente calificaban al curso como uno de los temas más importantes e interesantes de la especialización (y estamos hablando de un curso de estadística), sino también en el sentido de que nos encontramos años después con alumnos que nos dicen: "Yo estaba a años luz de los otros graduados gracias a su libro" o "aun cuando en la actualidad no realizo investigaciones, su curso realmente me ha ayudado en la lectura de las publicaciones científicas relacionadas con mi especialidad".

El reconocimiento a la primera edición ha sido sobrecogedor. Hemos recibido un gran número de e-mails y cartas de profesores (¡e incluso de alumnos!) agradeciéndonos desde todo el mundo de habla inglesa. Por supuesto, nos emocionó también la crítica entusiasta del *Contemporary Psychology** (Bourgeois 1997).

En la segunda edición hemos intentado mantener los aspectos del libro que fueron especialmente reconocidos, a la vez que trabajamos sobre el mismo para incluir aquellos otros aspectos surgidos de la respuesta de la gente, de nuestras propias experiencias, y de los avances y cambios en la materia. Sin embargo, antes de comenzar con la segunda edición quisiéramos reiterar algunos

* N. de la Trad.: *Psicología Contemporánea.*

comentarios realizados en la primera sobre la historia de los textos de esta especialidad y sobre aquellas cosas que hemos cambiado.

BREVE HISTORIA DE LOS TEXTOS DE ESTADÍSTICA COMO GÉNERO

En las décadas de 1950 y 1960, los textos sobre estadística eran libros aburridos, intimidatorios y basados esencialmente en la matemática, los cuales rezagaban rápidamente a la mayoría de los alumnos. En la década de 1970, se produjo una revolución; surgió el método intuitivo, con mucho menos énfasis en derivaciones, pruebas y fundamentos matemáticos. El nuevo método funcionó. Los alumnos comenzaron a perder el temor a los cursos de estadística y a considerarlos más accesibles e, incluso, bastante claros.

La tendencia intuitiva continuó en la década de 1980, y en la de 1990 se agregaron algunos trabajos realmente claros. En la actualidad, algunos textos además han comenzado a incentivar a los alumnos a utilizar las computadoras para realizar análisis estadísticos. Sin embargo, las exposiciones de interpretaciones intuitivas son cada vez más breves. Lo común es encontrar una especie de minimalismo en el que se hace una revisión superficial de la idea principal, y algunas veces se incluye la fórmula "de definición" de cada técnica. Después se detallan los procedimientos y ejemplos para la realización efectiva del cálculo, utilizando otra fórmula denominada "de cálculo".

Aun con toda esta modernización, o tal vez a causa de ella, al finalizar el curso, la mayoría de los alumnos no están en condiciones de explicar claramente la lógica implícita en las técnicas que han aprendido. Al transcurrir unos meses, difícilmente puedan realizar siquiera los procedimientos. Y lo que es más importante, no se cumplen los tres objetivos principales de los cursos de introducción a la estadística: a los alumnos les resulta imposible comprender el sentido de los resultados en las publicaciones de investigación psicológica, están mal preparados para futuros cursos de estadística (en los que los profesores deben, inevitablemente, dedicar la mitad del semestre para volver a enseñar el curso de introducción) y no se ha producido el contacto con el pensamiento profundo que, en teoría, justifica el cumplimiento del curso con las exigencias de la educación en general en el área cuantitativa.

¿QUÉ COSAS HEMOS REALIZADO DE MANERA DIFERENTE?

Continuamos haciendo lo que los mejores libros modernos ya están realizando: poner el acento en el aspecto intuitivo y quitárselo al aspecto matemático y, además, explicar cada tema en un lenguaje claro y simple. Pero nuestra obra se diferencia de esos otros libros en 11 puntos clave.

1. **Las fórmulas de definición retoman el centro de la escena,** ya que las mismas brindan un resumen simbólico conciso de la lógica de cada procedimiento en particular. Todas nuestras explicaciones, ejemplos, ejercicios e ítems de los paquetes de pruebas se basan en las mencionadas fórmulas de definición. (Hemos reducido adecuadamente las cifras utilizadas en los ejercicios y en los ítems de las pruebas para que los cálculos sean manejables).

¿Por qué utilizamos este método? Hasta el momento, los libros de estadística no han logrado amoldarse a la realidad tecnológica. Lo importante no es que los alumnos aprendan a calcular una prueba t con gran cantidad de números– ya que las computadoras pueden realizar ese trabajo. Lo importante es que los alumnos tengan siempre en mente la lógica implícita del procedimiento. Por ejemplo, analicemos la varianza poblacional, el promedio de los desvíos cuadráticos de la media. El concepto se presenta claramente a través de la fórmula de definición (una vez que el alumno se familiariza con los símbolos): varianza = $\sum (X - M)^2/N$. Al resolver una y otra vez los ejercicios utilizando esta fórmula, el significado de la misma se fija en la mente del alumno. Por el contrario, la versión de cálculo habitual de la mencionada fórmula sólo oscurece el significado: varianza = $[\sum X^2 - (\sum X)^2/N]/N$. ¡Lo único que se logra al resolver ejercicios utilizando esta segunda fórmula es enseñar al alumno la diferencia entre $\sum X^2$ y $(\sum X)^2$!

Enseñar fórmulas de cálculo en la actualidad es un anacronismo. Hoy en día, los investigadores realizan sus estadísticas con computadoras. Al mismo tiempo, el empleo de software estadístico hace que la comprensión de los principios básicos, tal como se expresan simbólicamente en las fórmulas de definición, sean más importantes que nunca.

El motivo por el cual los libros de estadística no han modificado sus métodos, con el advenimiento del software estadístico, es un misterio para nosotros, pero estamos convencidos de que el cambio ya debería haberse realizado. Por supuesto, dado que las fórmulas de cálculo son interesantes desde el punto de vista histórico y, ocasionalmente necesarias, y porque además algunos profesores pueden sentirse desprotegidos sin ellas, aun así las proporcionamos (con un ejemplo resuelto) en un breve apéndice de cada capítulo en el que normalmente se presentaría una fórmula de cálculo.

2. **Cada procedimiento es explicado tanto numérica como verbalmente y, en general, también en forma visual, describiendo los mismos ejemplos en cada una de las formas mencionadas.** Los ejercicios prácticos y los ítems de los paquetes de pruebas requieren, a su vez, que los alumnos calculen resultados, creen gráficos o ilustraciones, y redacten, además, una breve explicación del significado de las estadísticas en lenguaje lego. El material de cada capítulo, que incluye al menos dos ejemplos resueltos de las diversas formas mencionadas, prepara a los alumnos para los ejercicios y las preguntas de las pruebas.

La experiencia nos demuestra que las diferentes formas mencionadas para expresar una idea son de suma importancia para establecer un concepto de modo inalterable en la mente del alumno. Muchos estudiantes de psicología tienen mayor facilidad para manejarse con las palabras que con los números. En realidad, algunos tienen miedo de todo lo relacionado con la matemática. Al redactar las explicaciones en lenguaje lego tienen la oportunidad de hacer lo que mejor hacen y, si tienen dificultades, se ven forzados a enfrentarlas y a plantear los procedimientos en la forma verbal que mejor manejan.

3. **Hacemos hincapié en el hecho de que la estadística es un campo de investigación vivo y en crecimiento.** Dedicamos el tiempo necesario para describir las controversias y los desarrollos recientes en términos sencillos. El objetivo es que los alumnos tomen conciencia de que los métodos estadísticos son esfuerzos humanos destinados a dar sentido a grandes cantidades de datos; que las estadísticas no "surgen" por naturaleza, no son infalibles, ni son una descripción perfecta de los hechos que intentan describir, sino que constituyen un lenguaje en constante perfeccionamiento a través del pensamiento esmerado de aquellos que lo utilizan. Esperamos que esta orientación ayude al alumno a mantener una actitud inquisitiva y alerta como tal, como también a estar a la altura de los nuevos desarrollos estadísticos como profesional.

4. El objetivo principal de todo curso introductorio de estadística aplicada a la psicología es **preparar a los alumnos para la lectura de publicaciones científicas.** En realidad, el modo en que una publicación científica describe un procedimiento como la prueba *t* o el análisis de varianza con frecuencia es muy diferente de lo que el alumno espera encontrar, teniendo en cuenta las exposiciones que normalmente aparecen en los textos. Es por eso que este libro, a la vez que enseña un método estadístico, brinda ejemplos sobre el modo en que dicho método es presentado en las revistas científicas (extractos de publicaciones de actualidad). Los ejercicios prácticos y los ítems de los paquetes de pruebas también incluyen extractos de publicaciones para que los alumnos expliquen.

5. El libro está extraordinariamente **actualizado.** Por alguna razón, en la mayoría de los textos de introducción a la estadística que hemos visto, parecería que los autores estuvieran escribiendo en la década de 1950. Los principios básicos siguen siendo los mismos, pero las sutilezas con las que los estadísticos e investigadores analizan esos principios básicos han cambiado radicalmente. En la actualidad, los principios básicos están apuntalados por una apreciación diferente de temas, tales como la magnitud de efecto, la potencia y la acumulación de resultados por meta-análisis; la función primordial que desempeñan los diversos diseños; la coherencia implícita en las estadísticas por diferencia y por asociación, y la creciente prominencia de la regresión y los métodos relacionados con ella, al igual que un sinnúmero de nuevas orientaciones que surgen del papel preponderante de la computadora en el análisis. Estamos profundamente comprometidos con los últimos desarrollos con respecto a la teoría y aplicación de la estadística, y confiamos en que el libro refleja dicho compromiso. Por ejemplo, dedicamos todo un capítulo al tamaño de efecto y la potencia y, al discutir el manejo de situaciones en las que los supuestos no se cumplen, cubrimos el tema de las transformaciones de datos (se trata de un método ampliamente utilizado y de fácil comprensión para alumnos de nivel introductorio, pero que, sin embargo, la mayoría de los textos actuales de introducción a la materia rara vez menciona). Por supuesto, las secciones que tratan sobre controversias y desarrollos recientes son fundamentales para hacer de este libro un texto actualizado con respecto a la utilización efectiva de la estadística en la investigación actual.

6. **El capítulo 16** es único en cuanto **integra las técnicas más importantes que han sido enseñadas,** explicando que la prueba *t* es un caso especial del análisis de varianza y que tanto la prueba *t* como el análisis de varianza son casos especiales de correlación y regresión. (Resumiendo, presentamos el modelo lineal general). En el pasado, si este tema se trataba siquiera, sólo ocurría en textos avanzados. Sin embargo, muchos alumnos lo consideran valioso para asimilar y retener lo que han aprendido, así como también para sentir que han profundizado en los fundamentos de los métodos estadísticos.

7. **El último capítulo analiza los procedimientos avanzados** sin tratarlos en detalle. Explica en términos simples cómo interpretar esas estadísticas cuando aparecen en publicaciones científicas. La mayoría de las publicaciones científicas de psicología utilizan métodos tales como el análisis de covarianza, el análisis de covarianza multivariado, la regresión jerárquica múltiple, el análisis factorial y el modelo de ecuación estructural. Los alumnos que terminan un curso estándar de introducción a la estadística no cuentan con los elementos necesarios para comprender la mayoría de las publicaciones que deben leer para preparar sus trabajos o estudiar para determinado curso. El capítulo mencionado utiliza los principios básicos que los alumnos acaban de aprender (además de extractos extensivos de publicaciones científicas actuales) para lograr una interpretación rudimentaria de los procedimientos avanzados. A la vez, el capítulo sirve como guía que los alumnos pueden guardar y utilizar en el futuro al leer el tipo de publicaciones mencionadas.

8. El libro ha sido escrito con la intención de **apelar a las motivaciones que llevan a un alumno a especializarse en psicología.** Además de intentar representar la diversidad de la psicología, nuestros ejemplos destacan aquellos temas y poblaciones de mayor interés para los alumnos. El primer ejemplo fue extraído de un estudio real en el que 151 alumnos evalúan el nivel de estrés que sienten durante la primera semana de clases de un curso de introducción a la estadística. Otros ejemplos hacen hincapié en la psicología clínica, empresarial y educativa, y a la vez se incluyen suficientes e interesantes ejemplos de las áreas experimentales, sociales, de desarrollo y otras, que estimulan a los alumnos con el valor de dichas áreas. Además, en nuestros ejemplos destacamos continuamente la utilidad de los métodos estadísticos como herramientas para el proceso de investigación, evitando siempre que los alumnos sientan que lo que están aprendiendo es teoría por la teoría misma. El apéndice A brinda una visión general de métodos de investigación,

mostrando el contexto en el cual funciona la estadística. Y a medida que se enseña cada técnica se ilustra y recalca su función dentro del proceso de investigación.

9. La ***Guía de estudio y libro de tareas de computación para el alumno***, que acompaña esta obra, se concentra en el dominio de los conceptos, e incluye también instrucciones y ejemplos para resolver los ejercicios utilizando una computadora. La mayoría de las guías de estudio se concentran en la ubicación de números dentro de las fórmulas y en la memorización de reglas (coherentemente con el estilo de los textos que acompañan). Nuestra *Guía de estudio y libro de tareas de computación* establece, para cada capítulo, objetivos de aprendizaje, un resumen detallado del capítulo, las fórmulas tratadas en el capítulo correspondiente (con la definición de cada símbolo) y resúmenes de los pasos a seguir para la realización de cada proceso tratado en el capítulo, más una serie de exámenes para auto-calificarse que incluyen ejercicios de multiple-choice, ejercicios para completar y preguntas para contestar en forma de ejercicio o de ensayo. Además, para cada procedimiento tratado en el capítulo, la guía de estudio brinda pautas completas para la redacción de un ensayo, explicando el procedimiento a una persona que nunca ha asistido a un curso sobre estadística. Como una ayuda extra para el estudio, incluye tarjetas recortables de consulta rápida con todos los términos clave.

Es especialmente importante el hecho de que nuestra *Guía de estudio y libro de tareas de computación* brinda la ayuda necesaria para enseñar a los alumnos a realizar análisis estadísticos en una computadora. En primer lugar, hay un apéndice especial que presenta la terminología y los procedimientos del SPSS para Windows. Luego, en concordancia con los capítulos del libro, existe una sección que explica en forma detallada cómo realizar con una computadora los procesos tratados en cada capítulo. (Esta sección incluye instrucciones paso a paso, ejemplos e ilustraciones que muestran cómo se ve en la pantalla de la computadora cada ingreso y devolución de información). La guía cuenta también con actividades especiales para utilizar la computadora, con el fin de profundizar la comprensión de los temas. Hasta donde nuestro conocimiento nos ha permitido investigar, no existe otro paquete de textos sobre estadística que brinde tal flexibilidad o profundidad en el tratamiento de los diversos temas.

10. Hemos escrito también un *Manual para el Instructor* **que realmente ayuda a dictar el curso.** El manual comienza con un capítulo que resume lo que hemos percibido por nuestra propia experiencia en la enseñanza y el material producido por la investigación acerca de la efectividad en la enseñanza universitaria. El siguiente capítulo trata sobre organizaciones alternativas para el curso, que incluye posibles cronogramas y un programa a modo de ejemplo. Luego, cada capítulo, en concordancia con los capítulos del libro, brinda un resumen completo de la clase y **ejemplos resueltos que no se encuentran en el libro** (en un formato adecuado para realizar transparencias o para distribuir a los alumnos). Este material es particularmente útil, ya que crear ejemplos resueltos es una de las mayores dificultades al preparar clases sobre estadística.

11. Nuestro **"Banco de pruebas" y "Respuestas a los ejercicios de la serie II" facilita la preparación de buenos exámenes.** Para cada capítulo proporcionamos aproximadamente 40 ejercicios de multiple-choice, 25 ejercicios para completar y 10 ó 12 preguntas para contestar en forma de ejercicio o de ensayo. Teniendo en cuenta que el énfasis del curso está puesto en lo conceptual, los ejercicios de multiple-choice serán particularmente útiles para aquellos que no están preparados para calificar ensayos. Este suplemento también incluye las respuestas a la serie II de ejercicios de cada capítulo del libro, las cuales no aparecen en el mismo. (El libro incluye las respuestas a todos los ejercicios de la serie I y, al menos, un ensayo a modo de ejemplo, por cada capítulo).

FACTORES QUE INFLUYERON EN LA SEGUNDA EDICIÓN

La revisión para la segunda edición la realizamos en Manhattan. Esperamos que este hecho no haya provocado la pérdida de cualquier encanto que pudiera haber ganado la primera edición por haber sido realizada en París. Por otro lado, la presente edición está impregnada por el espíritu del teatro y del ballet.

Más aún, la presente revisión está enriquecida por la experiencia obtenida al enseñar con la anterior, y por la experiencia y aliento recibidos de infinidad de profesores que nos han escrito sobre sus propias experiencias al utilizar el libro.

La revisión también ha sido moldeada por nuestra propia aplicación de métodos estadísticos: los últimos cinco años han sido un período muy productivo para nosotros en nuestros propios programas de investigación, en el campo de la psicología social y de la personalidad. (Para tener una visión general de nuestros propios programas de investigación, remitirse a A. Aron & E. Aron, 1997; E. Aron & A. Aron, 1997). Tal vez haya sido especialmente útil que, durante los últimos tres años, uno de nosotros (AA) se haya desempeñado como editor asociado del JPSP (*Journal of Personality and Social Psychology, Revista Científica de Psicología Social y de la Personalidad)*, hecho que nos permite estar en contacto con el modo en que los mejores investigadores utilizan las estadísticas (como también con la forma en que los críticos califican el uso que sus colegas hacen de las mismas).

La revisión ha sido afectada también, y en gran medida, por los desarrollos ocurridos durante estos últimos cinco años en el campo de la estadística y en su aplicación a la psicología. Lo más importante en este sentido ha sido la controversia sobre el valor de las pruebas de significación y las correspondientes propuestas de reemplazarlas por las estimaciones puntuales del tamaño de efecto y los intervalos de confianza. Los temas que están surgiendo determinarán, sin duda, el modo en que los psicólogos utilicen la estadística en las próximas décadas. En el corto plazo, la influencia ha sido sorprendentemente escasa. Durante estos tres años como editor del JPSP, AA ha manejado aproximadamente 150 manuscritos y, sin embargo, sólo uno ha mostrado evidencia de la controversia actual. Editores de otras revistas científicas nos informan que sus experiencias son similares. Por lo tanto, consideramos que en el futuro cercano los alumnos continuarán necesitando estar bien familiarizados con la prueba de significación tradicional para estar en condiciones de leer publicaciones tanto nuevas como antiguas.

De todos modos, en esta revisión hemos intentado tener en cuenta los aparentes vientos de cambio. En especial, hemos mejorado nuestro tratamiento de los intervalos de confianza, así como también hemos cubierto más extensamente los temas involucrados en el debate actual sobre prueba de hipótesis (*véanse* las secciones de controversia de los capítulos 5, 7 y 8). Tal vez lo más importante sean los cambios sutiles de terminología en todo el texto, sobre la base de nuestros conocimientos de las cuestiones actuales. Nuestro objetivo es preparar alumnos para quienes este libro resulte útil ante cualquier cambio que se avecine, pero, al mismo tiempo, asegurarnos de que conocen los principios básicos tal como existen en la actualidad. Por eso, mientras todos están de acuerdo con que las pruebas de significación han sido mal utilizadas con demasiada frecuencia, nosotros recalcamos precisamente aquellos aspectos que permiten asegurarnos de que los alumnos no repetirán los errores más comunes.

CAMBIOS ESPECÍFICOS EN LA SEGUNDA EDICIÓN

Las revisiones que hemos realizado se pueden dividir en cuatro clases principales:

1. **Redacción:** hemos revisado cuidadosamente cada oración, simplificando las construcciones y la terminología siempre que fuera posible. Ya es bastante difícil aprender estadística como para tener que lidiar, además, con oraciones complicadas.

2. **Actualización de ejemplos:** hemos reemplazado más de 100 ejemplos de la primera edición con otros nuevos publicados durante los últimos dos años. Esta revisión es particularmente importante en las secciones acerca de cómo interpretar y evaluar las publicaciones científicas. La única finalidad de esas secciones es que los alumnos observen cómo se presentan las estadísticas cuando se informan en investigaciones de actualidad. Al revisar los viejos ejemplos y encontrar otros nuevos, nos sorprendió la existencia de no pocos cambios sutiles en el modo en que se informan los resultados estadísticos. Por ejemplo, hace cinco años, los efectos interactivos en el análisis de varianza se informaban, por lo general, con gráficos de líneas. En la actualidad, generalmente se utilizan gráficos de barras (*véase* el capítulo 13).

3. **Actualización del contenido y las controversias:** hemos actualizado el contenido teniendo en cuenta nuevos desarrollos en el campo relevantes para el curso de estadística básica. Incluye el material mencionado anteriormente sobre la controversia en cuanto a la prueba de hipótesis, al igual que varios otros cambios, tales como la influencia del escrito de DeCarlo de 1997 sobre curtosis o el de Frick de 1995 sobre prueba de la hipótesis nula. Las revisiones de contenido también tuvieron en cuenta cambios de terminología básicos, como puede ser la utilización del término "participantes" en lugar de "sujetos", conforme al estilo actual de la Asociación Americana de Psicología.

4. **Correcciones para mejorar la pedagogía y cumplir en mejor medida con las necesidades de los profesores que utilizan el libro:** hemos agregado nuevas secciones sobre probabilidad y análisis de varianza de medidas repetidas (*véanse* apéndices de los capítulo 5 y 13), una sección en el capítulo 1 sobre niveles de medición y una sección substancial sobre intervalos de confianza. Hemos escrito, letra por letra, la mayoría de los subíndices, e incluso nos hemos esforzado aún más que en la primera edición para utilizar ejemplos multiculturales siempre que fuera posible.

Algunos aspectos que no hemos cambiado. Los once puntos mencionados anteriormente en esta misma introducción continúan siendo las características centrales y distintivas del libro. Siempre que pudimos, evitamos también cambiar los ejemplos que incluían grandes tablas de cálculos para minimizar las posibilidades de error.

MANTENGÁMONOS EN CONTACTO

Es nuestro objetivo colaborar en todo lo que sea posible para que tenga éxito con su curso. Si usted tuviera alguna duda o sugerencia, por favor escríbanos o envíenos un e-mail (**aron@psych1.psy.sunysb.edu** es la dirección de ambos). Si, Dios no lo permita, usted encontrara un error en algún lugar del libro, prometemos que a) lo corregiremos en la siguiente edición, b) enviaremos los detalles a todos aquellos en la red y c) incluiremos su nombre en nuestros agradecimientos en el prefacio de la próxima edición.

AGRADECIMIENTOS

Ante todo, queremos agradecer a nuestros alumnos de todos estos años por haber dado forma a nuestro método de enseñanza, premiándonos con su valoración por las cosas que hemos hecho bien al igual que con sus diversas formas de anular lo que no hemos hecho tan bien.

Por habernos impulsado a iniciar este proyecto, queremos agradecer a nuestro amigo Eryan Strong, quien en primer lugar nos alentó para que lo emprendiéramos, y a Brete Harrison, quien guió el proyecto durante su desarrollo inicial. Agradecemos también la colaboración y apoyo de nuestro amigo John Touhey, quien leyó varios de los primeros borradores de capítulos. Los revisores del libro en diversas etapas han sido sumamente útiles identificando falencias en la lógica y la pedagogía, y sus elogios generosos nos dieron ímpetu cuando, ocasionalmente, nos sentíamos perdidos en la inmensidad del proyecto. Queremos agradecer a Paul C. Amrhein, Universidad de Nueva México; James V. Couch, Universidad James Madison; Livia M. D'Andrea, Universidad de Nevada, Reno; Susan E. Dutch, Universidad Estatal de Westfield; Peter C. Hill, Universidad de Grove City; J. Robert Newman, Universidad del Estado de California, Long Beach; Michael L. Frank, Universidad Estatal de Stockton; Martin A. Johnson, Universidad del Estado Occidental de Missouri; Carol Pandey, Universidad L. A. Pierce; Roger Bakeman, Universidad del Estado de Georgia; Jeffrey S. Berman, Universidad del Estado de Memphis; y Michael J. Scozzaro, Universidad SUNY en Buffalo.

RECONOCIMIENTOS

CO-1, PhotoDisc, Inc.; CO-2, David Young-Wolff/PhotoEdit; CO-3, Leonard Lee Rue, III/Photo Researchers; CO-4, Secretaría de Turismo de Nueva México; CO-5, H., Fouque/Photo Researchers, Inc.; CO-6, Len Rue, Jr./Photo Researchers; CO-7, Grant Heilman Photography; CO-8, PhotoDisc, Inc.; CO-9, U.S. Secretaría de Agricultura; CO-10, Chip Henderson Photography; CO-11, Tom Hollyman/Photo Researchers, Inc.; CO-12, Bill Bachman/Photo Researchers, Inc.; CO-13, Okoniewski/The Image Works; CO-14, Barry L. Runk/Grant Heilman Photography; CO-15, Michael Newman/PhotoEdit; CO-16, Matura/Gamma-Liaison, Inc.; CO-17, Simon Fraser/Science Photo Library.

Introducción para el alumno

El objetivo del libro es ayudar a **comprender** la estadística concentrándose en los significados y conceptos, no sólo en los símbolos y números.

Esto juega a favor del alumno. La mayoría de aquellos que se especializan en psicología no aman los cálculos, pero sí se sienten sumamente cómodos con las ideas; por lo tanto, queremos subrayar lo siguiente según nuestra experiencia de 30 años de enseñanza: **nunca hemos tenido un alumno que, siendo capaz de tener un buen desempeño en otras materias universitarias, no lo haya logrado en esta materia.** (No obstante, debemos admitir que tener éxito en esta materia puede requerir más esfuerzo que tenerlo en las otras).

En esta introducción trataremos los motivos por los que se estudia la materia y cómo aprovecharla al máximo.

¿POR QUÉ APRENDER ESTADÍSTICA? (ADEMÁS DE CUMPLIR CON UN REQUISITO)

1. Entender la estadística es crucial para poder leer publicaciones científicas de psicología. Casi todos los cursos que el alumno tome como especialista en psicología harán hincapié en los resultados de estudios científicos, y estos usualmente se expresan a través de las estadísticas. Si el alumno no comprende la lógica básica de las estadísticas, si no puede comprender la jerga, las tablas y los gráficos que constituyen el centro de cualquier informe científico, la lectura de los resultados de investigaciones científicas será muy superficial.

2. **Comprender la estadística es crucial para poder realizar investigaciones científicas.** Con el tiempo, muchos especialistas en psicología deciden realizar estudios de posgrado. Los estudios de posgrado en psicología, incluso en psicología clínica o de asesoramiento y otras áreas aplicadas, casi siempre incluyen la **realización** de investigaciones científicas. Con frecuencia, aprender a realizar investigaciones es el punto central de los estudios de posgrado, y realizar investigaciones científicas casi siempre incluye el empleo de estadísticas. Este curso proporciona

al alumno una base sólida de los conocimientos estadísticos necesarios para realizar investigaciones. Más aún, dominando la lógica básica y el razonamiento estadístico, el alumno estará extraordinariamente bien preparado para cursos avanzados que se concentran en la parte más concreta del análisis de investigaciones científicas.

Muchos programas de psicología ofrecen también oportunidades de realizar investigaciones científicas a alumnos no graduados. La idea principal de este libro es que el alumno comprenda la estadística, no que la utilice. Aun así, el alumno aprenderá lo necesario para realizar los análisis más básicos de información utilizados en el tipo de investigaciones que probablemente realice.

3. **Comprender la estadística desarrolla el pensamiento analítico y crítico del alumno.** Los especialistas en psicología con frecuencia están interesados principalmente en las personas y en mejorar el mundo real. Lo dicho anteriormente no significa que los especialistas en psicología eviten las ideas abstractas, de hecho, a los alumnos que conocemos los estimulan los niveles de abstracción prácticamente filosóficos en los que muchas veces parecen esconderse los secretos de las experiencias humanas. Sin embargo, incluso este tipo de ideas abstractas, por lo general, al principio sólo se captan superficialmente como frases hechas en lugar de conocimientos útiles. De todas las materias que el alumno posiblemente estudie en la carrera de psicología, es probable que ésta sea la que más lo ayude a aprender a pensar en forma precisa, a evaluar información y a aplicar el análisis lógico a alto nivel.

CÓMO APROVECHAR EL CURSO AL MÁXIMO

Al respecto, podemos ofrecer cinco consejos:

1. **Concéntrese en los conceptos.** Considere este curso no tanto como un curso de matemática sino como uno de lógica. Cuando lea una sección de un capítulo, concentre su atención en captar los principios. Al realizar los ejercicios piense en las razones por las que realiza cada paso; si intenta simplemente memorizar cómo obtener los números correctos, habrá aprendido muy poco que le pueda ser útil en sus estudios futuros, y tampoco tendrá muy buenos resultados en los exámenes de este curso.

2. **Asegúrese de comprender cada concepto antes de pasar al siguiente.** La estadística es acumulativa. Cada nuevo concepto se construye sobre el anterior. Incluso dentro de un mismo capítulo, si ha leído una sección y no la comprende, **deténgase**, vuelva a leerla, a razonarla y pida ayuda. Es preciso que realice todo lo necesario para captar el significado. (Si considera que ha comprendido una sección, pero no está totalmente seguro, intente realizar un ejercicio pertinente de los que aparecen al final del capítulo).

Tener que leer el material del libro una y otra vez no significa que uno carezca de capacidad. La mayoría de los alumnos necesitan leer cada capítulo varias veces, y cada lectura es mucho más lenta que en el caso de un texto común. La lectura de textos de estadística debe realizarse detenidamente, con calma y concentración, para que vaya surgiendo el significado. Es importante dedicar mucho tiempo a este tipo de lectura, así como también a su relectura.

3. **No se retrase.** Debido a que la estadística es acumulativa, si se retrasa en la lectura o pierde clases, las clases a las que luego asista resultarán prácticamente incomprensibles. Y cada vez será más difícil ponerse al día.

4. **Estudie con especial intensidad durante la primera mitad del curso.** Es particularmente importante dominar completamente los temas tratados al comienzo del curso, ya que todo lo demás en estadística se construye sobre los primeros conceptos aprendidos. Sin embargo, comúnmente el comienzo del semestre es la época en que los alumnos estudian menos seriamente.

Si ha logrado dominar la primera mitad del curso– no sólo aprender la idea general sino conocerla realmente–, la segunda mitad resultará más sencilla. Si no ha logrado dominar la primera parte, la segunda resultará casi imposible.

5. **Ayúdense entre ustedes.** No existe mejor forma de afianzar y profundizar los conocimientos de estadística que intentar explicarlos a alguien a quien le resulta más complicado aprenderlos. (Por supuesto, esto debe hacerse con paciencia y respeto). Para aquellos a los que la materia les resulta más difícil, no existe mejor forma de resolver las dificultades que aprender de otro alumno que acaba de comprender a fondo esos temas.

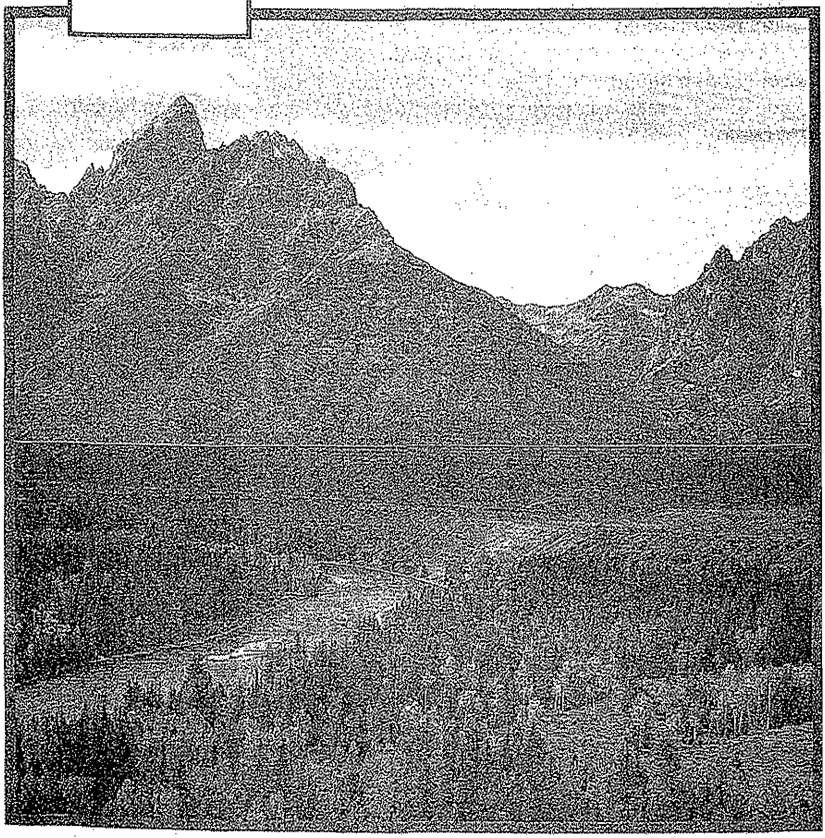
Por eso, recomendamos firmemente que se formen grupos de estudio de dos a cuatro alumnos. Lo ideal sería que los grupos incluyan alumnos que esperan entender el material fácilmente y otros que no. Aquellos que comprendan la estadística con facilidad se beneficiarán ayudando a otros a los que les cueste más; éstos últimos pondrán a prueba enormemente los supuestos conocimientos de los primeros. Aquellos que crean que tendrán inconvenientes, necesitan trabajar con los que no los tienen; que un ciego guíe a otro ciego no es una buena forma de aprender. También es conveniente escoger compañeros de estudio que vivan cerca para que sea fácil reunirse, y también hacerlo frecuentemente, si es posible una vez entre clase y clase.

COMENTARIO FINAL

Aunque cueste creerlo, nos encanta enseñar estadística. Una y otra vez hemos vivido la maravillosa experiencia de que se nos acerquen alumnos rebosantes de alegría a decirnos: "Profesor Aron, saqué un 90 en este examen, ¡no lo puedo creer! ¡Yo, un 90 en un examen de estadística!" O el alumno que nos confiesa: "Realmente es entretenido. No se lo diga a nadie, pero en verdad me divierte la estadística", ¡nada menos! Esperamos que a usted le ocurra algo parecido en este curso.

Arthur Aron
Elaine N. Aron

1 Organización de un grupo de números



Descripción del capítulo

- Las dos ramas de la metodología estadística.
- Tablas de frecuencias.
- ¿Cómo crear una tabla de frecuencias?
- Tabla de frecuencias agrupadas.
- Histogramas.
- Polígonos de frecuencias.
- Formas de las distribuciones de frecuencias.
- Controversias y limitaciones.
- Tablas de frecuencias, histogramas y polígonos de frecuencias según se describen en publicaciones científicas.
- Resumen.
- Términos clave.
- Ejercicios.

Queremos darle al lector la bienvenida a nuestro libro de estadística. Imaginamos que se asemeja a otros estudiantes de psicología que hemos conocido: eligió cursar esta materia porque le fascinan las personas, sus comportamientos visibles, y tal vez también su vida interior e incluso la propia. Algunos lectores son altamente científicos; otros más intuitivos. A algunos les gusta la matemática, a otros no tanto; y algunos hasta le temen. Cualquiera sea la categoría en la que se encuentre el lector, es bienvenido y puede estar seguro de que si le presta especial atención a nuestro libro (tal vez un poco más que a la mayoría de los libros de texto), realmente aprenderá estadística. El método utilizado en este libro resultó de gran utilidad en la enseñanza de la materia a todo tipo de alumnos, incluso a aquellos que previamente habían cursado estadística con resultados insatisfactorios. Estamos seguros de que con nuestro libro y la ayuda de un profesor aprenderá bien la estadística.

Lo más importante es que el lector sepa que no importa por qué razón estudia psicología. Este curso no pretende ser una pérdida de tiempo. La utilidad de esta materia radica en la necesidad de comprender la estadística para leer los trabajos realizados por otros psicólogos; también para realizar sus propias investigaciones y para pulir tanto su capacidad de razonamiento como su intuición. ¿Qué es realmente la estadística? Es una herramienta que ha evolucionado a partir de un proceso básico de pensamiento que todo psicólogo, todo ser humano, emplea: observamos algo; nos preguntamos cuál es su significado o cuál es su causa; aplicamos nuestra capacidad de discernimiento o nuestra intuición; volvemos a observar, pero ahora en detalle, o bien intentamos realizar algunos pequeños cambios en el proceso para probar nuestra intuición. Entonces nos enfrentamos al eterno problema: ¿Se confirmó o no nuestro presentimiento? ¿Cuáles son las posibilidades de que lo que hemos observado en esta segunda oportunidad suceda una y otra vez, de tal forma que podamos anunciar el resultado de nuestro razonamiento al mundo como algo probablemente cierto?

En otras palabras, la estadística es un método de búsqueda de la verdad. O al menos puede indicarnos las probabilidades de que nuestro presentimiento sea verdadero en este momento y en este lugar, con este tipo de personas. Esa búsqueda de la verdad, o al menos de la probabilidad futura, es la esencia de la psicología, de la ciencia y de la evolución humana. Pensemos en las primeras hipótesis: ¿Qué harán los mamuts la próxima primavera?; ¿qué sucederá si como esta raíz? Es fácil ver cómo han sobrevivido aquellos que han acertado, y el propio lector es uno de ellos. La estadística es una forma de búsqueda de precisión y verdad.

Los psicólogos utilizan métodos estadísticos para dar sentido a los números que reúnen al investigar. El problema de cómo diseñar una investigación adecuada es todo un tema en sí mismo, el cual resumiremos en el apéndice A. No obstante, en este libro nos limitamos a tratar los métodos estadísticos que dan sentido a los datos recolectados durante una investigación.

LAS DOS RAMAS DE LA ESTADÍSTICA

Existen dos ramas principales de la estadística:

1. **Estadística descriptiva**: los psicólogos la utilizan para resumir y hacer comprensibles los datos recolectados en el transcurso de una investigación.
2. **Estadística inferencial**: los psicólogos la utilizan para sacar conclusiones que, basadas en los datos recolectados durante la investigación, tienen una significancia que va más allá de los mismos.

En este capítulo y en los próximos tres, nos concentraremos en la estadística descriptiva. El tema es importante en sí mismo y, además, prepara al alumno para comprender la estadística inferencial, que constituye el tema central del resto del libro.

En este capítulo veremos cómo utilizar tablas y gráficos para describir grupos de datos. El propósito de la estadística descriptiva es facilitar la comprensión de esos datos, siendo las tablas y los gráficos muy útiles en ese aspecto.

TABLAS DE FRECUENCIAS

Comencemos con un ejemplo. Durante la primera semana del curso, Aron, Paris y Aron (1995), como parte de un estudio más amplio, repartieron un cuestionario a 151 alumnos en una clase de introducción a la estadística. Una de las preguntas era la siguiente: "¿Qué grado de estrés has experimentado en las últimas dos semanas y media, en una escala del 0 al 10, en la que 0 indica **para nada estresado,** y 10 **tan estresado como es posible**?" Las puntuaciones dadas por 151 estudiantes fueron las siguientes:

4, 7, 7, 7, 8, 8, 7, 8, 9, 4, 7, 3, 6, 9, 10, 5, 7, 10, 6, 8, 7, 8, 7, 8, 7, 4, 5, 10, 10, 0, 9, 8, 3, 7, 9, 7, 9, 5, 8, 5, 0, 4, 6, 6, 7, 5, 3, 2, 8, 5, 10, 9, 10, 6, 4, 8, 8, 8, 4, 8, 7, 3, 8, 8, 8, 8, 7, 9, 7, 5, 6, 3, 4, 8, 7, 5, 7, 3, 3, 6, 5, 7, 5, 7, 8, 8, 7, 10, 5, 4, 3, 7, 6, 3, 9, 7, 8, 5, 7, 9, 9, 3, 1, 8, 6, 6, 4, 8, 5, 10, 4, 8, 10, 5, 5, 4, 9, 4, 7, 7, 7, 6, 6, 4, 4, 4, 9, 7, 10, 4, 7, 5, 10, 7, 9, 2, 7, 5, 9, 10, 3, 7, 2, 5, 9, 8, 10, 10, 6, 8, 3

Tabla 1-1. Cantidad de individuos que elige cada valor de la escala de medición de estrés.

Puntuación	Frecuencia
10	14
9	15
8	26
7	31
6	13
5	18
4	16
3	12
2	3
1	1
0	2

Fuente: Aron, Paris & Aron (1995).

El sólo hecho de leer todas estas clasificaciones llevaría un tiempo considerable. Al examinar rápidamente los datos obtenemos una idea de la tendencia general, pero difícilmente sea un método preciso. Una solución es confeccionar una tabla que muestre cuántos alumnos eligieron cada uno de los once valores de estas puntuaciones (0, 1, 2 y siguientes, hasta 10). Es precisamente lo que hemos hecho en la tabla 1-1. Este tipo de tabla se denomina **tabla de frecuencia,** porque muestra con qué frecuencia (cuántas veces) ocurre cada puntuación. Una tabla de frecuencias hace que el patrón numérico se comprenda claramente y a simple vista. En este ejemplo, podemos ver que la mayoría de los alumnos se atribuyeron un nivel de estrés en alrededor de 7 u 8 puntos, y que muy pocos lo hicieron por debajo de esos valores.

Cuadro 1-1. Trivialidades importantes para estudiantes de estadística con espíritu poético.

La palabra **estadística** deriva de la palabra italiana *statista*, persona que trata asuntos de Estado (de *Stato*, "Estado"). Originalmente se la llamó "aritmética de Estado" e involucraba representar con tablas la información relativa a las naciones, especialmente aquellos datos relacionados con los impuestos y la planificación de la viabilidad de las guerras.

La estadística deriva de una amplia variedad de fuentes. La idea de recolectar estadísticas derivó de requerimientos gubernamentales, pero también de la necesidad, en tiempos antiguos, de calcular la posibilidad de naufragios y piratería con el propósito de administrar los seguros marítimos para fomentar viajes comerciales y de exploración a sitios lejanos. El estudio moderno de los índices de mortalidad y seguros de vida se originó en las fosas donde se depositaban los cadáveres de las víctimas de la plaga del siglo XVII; allí se contaban los cuerpos de personas muertas en el esplendor de su juventud. La teoría de errores (tratada en el capítulo 4 de este libro) se originó con la astronomía, en la observación de las estrellas; la teoría de la correlación (capítulo 3), con la biología, a partir de la observación de padres e hijos. La teoría de la probabilidad (capítulo 5) llegó a nosotros desde los tensos ambientes de las mesas de juego. La teoría del análisis de experimentos (capítulos 9 a 13) comenzó a desarrollarse en las destilerías y en los ondulantes campos de trigo, donde las predicciones correctas podían determinar no sólo la fabricación de una deliciosa cerveza sino también la supervivencia de miles de granjeros. Las teorías de la medición y el análisis factorial (capítulo 17) tienen su origen en la psicología de la personalidad, campo en el que por primera vez se exploraron las profundidades del carácter humano mediante la utilización de números. Y el chi cuadrado (capítulo 14) llegó a nosotros desde la sociología, que con frecuencia trata con clases sociales.

En los comienzos del desarrollo de la estadística, en los siglos XVII y XVIII, era usual que se utilizaran los nuevos métodos para probar la existencia de Dios. Por ejemplo, John Arbuthnot descubrió que en Londres, entre los años 1629 y 1970, nacieron más bebés de sexo masculino que femenino. Mediante lo que se considera el primer caso de utilización de una prueba estadística, probó que el índice de natalidad masculino era mayor de lo que la razón por azar hubiera indicado (asumiendo en éste un porcentaje del 50% para cada sexo), llegando a la conclusión de que se estaba cumpliendo un plan determinado para contrarrestar el hecho de que los hombres enfrentaban mayores peligros para obtener el sustento para sus familias, y que dicha planificación, según él, sólo podía haber sido realizada por Dios.

En el año 1767, John Michell también utilizó la teoría de la probabilidad para demostrar la existencia de Dios, cuando argumentó que las posibilidades de que seis estrellas se ubicaran tan cerca como lo estaban las de la constelación de Pléyades eran de 500.000 a 1, y que por ende su ubicación tenía que haber sido un acto deliberado del Creador.

La estadística ayudó a ganar las guerras independentistas de lo que luego serían los Estados Unidos de Norteamérica. John Adams obtuvo ayuda vital de Holanda después de tomar en cuenta ciertas estadísticas de suma importancia, cuidadosamente recolectadas por los clérigos en las parroquias locales. Las estadísticas demostraban que las colonias habían duplicado su población cada 18 años, agregando 20.000 hombres por año para la lucha. "¿Es éste el caso de nuestro enemigo, Gran Bretaña?", escribió Adams. "¿Entonces quién podrá mantener la guerra por más tiempo?".

En el año 1786, el presidente de los Estados Unidos, Thomas Jefferson, tuvo en cuenta estadísticas similares. Escribió que su pueblo "se intranquilizaba" cuando había más de diez habitantes por milla cuadrada y que, debido al crecimiento de la población del nuevo país, en cuarenta años estas almas inquietas llenarían todo el territorio disponible. Unos diecisiete años después, Jefferson duplicó el tamaño de ese territorio disponible a través de la adquisición de Louisiana.

En la actualidad, la estadística, en el sentido de "aritmética de Estado", es respaldada jurídicamente por la mayoría de los gobiernos. Por ejemplo, el primer artículo de la constitución de los Estados Unidos exige la realización de un censo.

¿Quién dijo que la estadística carece de alma o sentido humano?

Variables, valores y observaciones.

Antes de continuar con nuestra exposición sobre tablas de frecuencias, es necesario presentar cierta terminología que se aplica a esas tablas y prácticamente a todos los temas tratados en este libro. En primer lugar, explicaremos el significado de **variables**, **valores** y **observaciones**. Luego, consideraremos brevemente otro tema relacionado con diferentes tipos de variables: los niveles de medición.

Una manera de describir la función de una tabla de frecuencias es diciendo que muestra la frecuencia de cada uno de los **valores** de una determinada **variable**. Un valor es simplemente un número, como por ejemplo 4, -81 ó 367,12. Un valor también puede ser una categoría, como por ejemplo, masculino o femenino o la religión de una persona.

Una variable es una característica que puede tener diferentes valores. En otras palabras, puede **variar**. En el ejemplo que ofrecimos anteriormente, la variable es el nivel de estrés, con valores que van desde el 0 al 10. La estatura es una variable; la clase social es una variable; el resultado obtenido en una prueba de creatividad es una variable; el tipo de psicoterapia recibida por los pacientes es una variable; la velocidad en una prueba de tiempo de reacción es una variable; la cantidad de personas ausentes en el trabajo es una variable, y así sucesivamente.

En toda variable, cada persona analizada presenta un número o **valor observado** (observación) particular, que constituye el valor de esa persona con respecto a la variable. Por ejemplo, el valor observado de Chris en la variable de estrés podría haber sido un valor de 6; Pat podría tener un valor observado de 8. Con frecuencia utilizamos las palabras **valor observado** u observación para referirnos al valor particular de una persona en una variable, ya que la mayoría de las investigaciones psicológicas incluyen observaciones registradas en algún tipo de prueba.

Las investigaciones psicológicas no se tratan de otra cosa más que de variables, valores y observaciones. Utilizaremos estos términos a lo largo de todo el libro. Aunque las definiciones formales son un poco abstractas, en la práctica el significado de estos términos es generalmente obvio.

Niveles de medición: variables numéricas y nominales

La mayoría de las variables utilizadas por psicólogos son semejantes a las que aparecen en el ejemplo de las puntuaciones de estrés. Las observaciones son números que indican el grado o la cantidad de lo que se está midiendo. En el ejemplo del estrés, cuánto mayor era el número, mayor era el estrés. Nosotros nos referimos a este tipo de variables como **variables numéricas**. Las variables numéricas también se denominan **variables cuantitativas**.

En realidad, existen distintos tipos de variables numéricas. En las investigaciones psicológicas, la distinción más importante debe hacerse entre (a) variables en las que los números representan cantidades prácticamente iguales de aquello que se está midiendo y (b) variables en las que los números sólo representan posiciones relativas. Por ejemplo, el GPA (*Grade Pont Average*, promedio de calificaciones) es, en líneas generales, una **variable intervalar**, ya que la diferencia entre una calificación promedio de 2,5 y otra de 2,8 significa prácticamente lo mismo que la diferencia entre calificaciones de 3,0 y 3,3 (en ambos casos existe una diferencia de 0,3 puntos entre las calificaciones promedio).

Un ejemplo de **variable** sería la jerarquía en una clase determinada. La diferencia entre el segundo y el tercero de la clase, en cuanto al GPA, puede implicar una diferencia distinta de la que existe entre el octavo y el noveno. De alguna manera, se brinda menos información con una variable ordinal; ésta es menos precisa.

Otro tipo importante de variable en la investigación psicológica es la **variable nominal** (también llamada **variable categórica**). Las variables nominales son variables tales como el sexo o el diagnóstico psiquiátrico, es decir, aquellas en las que los valores son nombres o categorías (el término **nominal** proviene de la idea de que sus valores son nombres). Por ejemplo, los valores correspondientes al sexo son "femenino" y "masculino". La "observación" de cada persona en la variable sexo será uno de estos dos valores. Del mismo modo, el diagnóstico tiene valores tales como trastorno de estrés postraumático, esquizofrenia y trastorno de obsesión compulsiva.

Las distintas clases de variables mencionadas reflejan diferentes **niveles de medición**. Supongamos que un investigador está analizando los efectos de un tipo particular de lesión cerebral que incide en la capacidad de reconocimiento de objetos. Uno de los enfoques que podría aplicar el investigador sería medir la cantidad de objetos diferentes que una persona que padece una lesión puede observar al mismo tiempo. En ese caso estaríamos frente a un ejemplo de nivel de medición intervalar. Otra alternativa sería que el investigador calificara a las personas del siguiente modo: incapaces de observar objeto alguno (0); capaces de observar sólo un objeto a la vez (1); capaces de observar un objeto, con una vaga sensación de la existencia de otros objetos (2), o visión normal (3). En este caso, estaríamos frente a una medición ordinal. Finalmente, el investigador podría dividir a las personas entre aquellas que son completamente ciegas (B), aquellas que pueden identificar la ubicación de un objeto pero no pueden definir qué es ese objeto (L), aquellas que pueden identificar qué es el objeto pero no pueden localizarlo en el espacio (I), aquellos que pueden localizar e identificar un objeto pero sufren otras anormalidades en cuanto a la percepción de objetos (O), y aquellos con percepción visual normal (N). En este caso estaríamos frente a un nivel de medición nominal.

A lo largo de todo el libro, y como sucede en la gran mayoría de las situaciones reales de investigación psicológica, trabajamos con variables numéricas. Asimismo, trabajaremos principalmente con variables intervalares (o variables que se consideran bastante aproximadas a las mencionadas). En el capítulo 14, veremos métodos estadísticos que incluyen variables nominales, y en el capítulo 15, métodos relacionados con variables ordinales. No obstante, cabe destacar en este punto, que a menudo se hacen tablas de frecuencias con variables nominales. En ese caso, la tabla muestra la frecuencia de cada uno de los valores de la variable nominal. Por ejemplo, un psicólogo especializado en desarrollo, que analiza los estilos de disciplina utilizados por padres solteros, podría confeccionar una tabla de frecuencias indicando la cantidad de padres que utilizan cada uno de los siete estilos diferentes de disciplina. Sin embargo, en líneas generales, este libro se concentra principalmente en tablas de frecuencias y en otros procedimientos con variables numéricas intervalares.

¿CÓMO CONFECCIONAR UNA TABLA DE FRECUENCIAS?

Ahora podemos dedicarnos a los procedimientos concretos para la creación de una tabla de frecuencias.

Los pasos a seguir son tres:

1. Preparar una lista con cada valor posible, comenzando con el mayor y finalizando con el menor. En los resultados de la medición de estrés la lista va desde el 10, la mayor puntuación posible, hasta el 0, la menor puntuación posible.[1] (Aun cuando alguna de las puntuaciones entre 10 y 0 no se utilice, ese valor de la variable de estrés deberá ser incluido en la lista, mostrando que presenta una frecuencia 0. Por ejemplo, si nadie hubiera seleccionado un valor 2 de estrés, de todos modos debería incluirse el 2 como uno de los valores de la tabla de frecuencias).

2. Controlar una por una todas las observaciones, realizando una marca en cada una de ellas al lado del valor correspondiente. La figura 1-1 ilustra este procedimiento. Es recomendable ir tachando cada registro a medida que se incluya la marca correspondiente junto a la lista de valores.

3. Preparar una tabla prolija que muestre cuántas veces ha sido elegido cada valor de la lista. Para ello se debe sumar la cantidad de marcas realizadas junto a cada valor. Es conveniente controlar la precisión del trabajo realizado sumando esos totales, de modo de asegurarse de que coincidan con la cantidad total de observaciones. (*Véase* figura 1-1).

Ejemplo de la confección de una tabla de frecuencias

Como parte de un estudio más amplio sobre el comportamiento social de alumnos universitarios, Tracy McLaughlin-Volpe y sus colegas (1998) hicieron que 94 estudiantes del ciclo de Introducción a la Psicología llevaran un diario de sus interacciones sociales durante una semana del semestre. Cada vez que los participantes tuvieran una interacción social de 10 minutos o más, deberían llenar una tarjeta. La tarjeta incluía preguntas tales como quiénes eran las otras personas con las que interactuaban, cómo se sintió el alumno durante la interacción y varios aspectos rela-

[1] La mayoría de los expertos en estadística siguen el procedimiento aquí recomendado, ordenando los valores desde el mayor, en la parte superior, hasta el menor, en la parte inferior. Sin embargo, en las publicaciones científicas es más probable que las tablas de frecuencias contengan el número menor en la parte superior y el mayor en la parte inferior.

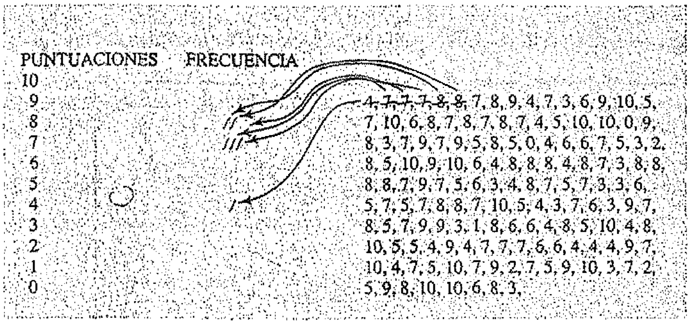
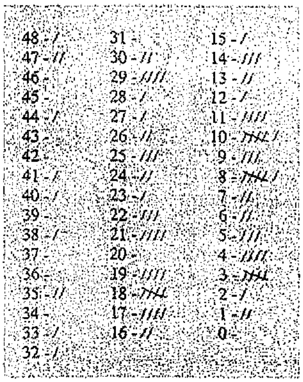


Figura 1-1. Construcción de una tabla de frecuencias, utilizando los datos de: Aron, Paris, & Aron (1995).

cionados con la naturaleza de la conversación mantenida. Excluyendo las situaciones familiares y laborales, la cantidad de interacciones sociales (de 10 minutos o más de duración) ocurridas durante una semana en el caso de cada uno de los 94 alumnos, fue la siguiente:



48, 15, 33, 3, 21, 19, 17, 16, 44, 25, 30, 3, 5, 9, 35, 32, 26, 13, 14, 14, 47, 47, 29, 18, 11, 5, 19, 24, 17, 6, 25, 8, 18, 29, 1, 18, 22, 3, 22, 29, 2, 6, 10, 29, 10, 21, 38, 41, 16, 17, 8, 40, 8, 10, 18, 7, 4, 4, 8, 11, 3, 23, 10, 19, 21, 13, 12, 10, 4, 17, 11, 21, 9, 8, 7, 5, 3, 22, 14, 25, 4, 11, 10, 18, 1, 28, 27, 19, 24, 35, 9, 30, 8, 26

Figura 1-2. Construcción de una tabla de frecuencias del número de interacciones sociales mantenidas por estudiantes durante una semana.
(**Fuente:** McLaughlin-Volpe et al. 1998).

Sigamos ahora los tres pasos indicados para la creación de una tabla de frecuencias.

1. Preparar una lista de cada valor posible a lo largo del margen izquierdo de la página, comenzando con el mayor y finalizando con el menor. En este estudio en particular, la mayor cantidad de interacciones podría ser cualquier número. Sin embargo, el mayor número obtenido en el grupo analizado es 48, por lo tanto, podemos utilizarlo como el mayor valor posible. Y la menor cantidad posible de interacciones es 0. Teniendo en cuenta lo anterior, el primer paso a seguir es hacer una lista de esos valores. (Sería buena idea utilizar varias columnas para poder incluir todos los registros en una sola página).

2. Revisar los registros uno por uno, haciendo una marca por cada uno junto al valor correspondiente en la lista. La figura 1-2 muestra el resultado de este paso.

3. Preparar una tabla prolija que indique cuántas veces ocurre cada uno de los valores de la lista. La tabla 1-2 es la tabla de frecuencias definitiva.

Tabla 1-2.
Tabla de frecuencias del número de interacciones sociales mantenidas por 94 estudiantes universitarios durante una semana.

Observaciones	Frecuencias	Observaciones	Frecuencias	Observaciones	Frecuencias
48	1	31	0	15	1
47	2	30	2	14	3
46	0	29	4	13	2
45	0	28	1	12	1
44	1	27	1	11	4
43	0	26	2	10	6
42	0	25	3	9	3
41	1	24	2	8	6
40	1	23	1	7	2
39	0	22	3	6	2
38	1	21	4	5	3
37	0	20	0	4	4
36	0	19	4	3	5
35	2	18	5	2	1
34	0	17	4	1	2
33	1	16	2	0	0
32	1				

Fuente: McLaughlin-Volpe et al. (1998).

Tabla de frecuencias agrupadas

A veces existen tantos valores posibles que es difícil reflejarlos en una tabla de frecuencias; en el último ejemplo sucedía algo así. La solución a este problema consiste en formar grupos de valores que incluyan todos aquellos valores que se encuentran comprendidos dentro de un determinado **intervalo**. Analicemos el ejemplo de medición del estrés. En lugar de presentar una frecuencia para los alumnos que atribuyeron a su estrés 8 puntos y otra para los que lo hicieron con un 9, podríamos presentar una categoría combinada de 8 y 9. La categoría combinada es un intervalo que incluye los dos valores, la cual tendría una frecuencia de 41 (26 casos con 8 puntos, más 15 casos con 9 puntos).

Una tabla que asocia frecuencias a intervalos es una **tabla de frecuencias agrupadas**. La tabla 1-3 es la tabla de frecuencias agrupadas correspondiente al ejemplo de medición del estrés. (Sin embargo, en este caso, la tabla de frecuencias completa tenía sólo 11 valores diferentes y, por ende, no era realmente necesario realizar una tabla de frecuencias agrupadas). La tabla 1-4 es la tabla de frecuencias agrupadas correspondiente al caso de la cantidad de interacciones sociales mantenidas por 94 estudiantes durante una semana.

Tabla 1-3.
Tabla de frecuencias agrupadas correspondiente a las puntuaciones de estrés.

Intervalos de clase:	Frecuencia
10–11,9	14
8–9,9	41
6–7,9	44
4–5,9	34
2–3,9	15
0–1,9	3

Fuente: Aron, Paris & Aron (1995).

Tabla 1-4.
Tabla de frecuencias agrupadas del número de interacciones sociales mantenidas por 94 estudiantes universitarios durante una semana.

Intervalos de clase	Frecuencia
45–49,9	3
40–44,9	3
35–39,9	3
30–34,9	4
25–29,9	11
20–24,9	10
15–19,9	16
10–14,9	16
5– 9,9	16
0– 4,9	12

Fuente: McLaughlin-Volpe et al. (1998).

A través de una tabla de frecuencias agrupadas se puede brindar información de forma aún más directa que mediante una tabla de frecuencias común. Cabe destacar, sin embargo, que la facilidad de comprensión que brinda una tabla de frecuencias agrupadas se logra a costa de perder cierta información, como por ejemplo, el detalle de las frecuencias dentro de cada intervalo.

¿Cómo confeccionar una tabla de frecuencias agrupadas?

El punto clave al construir una tabla de frecuencias agrupadas es definir de qué modo se agruparán los valores individuales, es decir, la serie de valores a incluir en cada intervalo. No deben existir brechas entre los intervalos, porque lo que nos interesa es incluir todos los valores de la tabla de frecuencias; tampoco deben superponerse unos valores con otros, porque no sabríamos en qué intervalo incluir algunas observaciones. Asimismo, es muy importante que todos los intervalos tengan la misma longitud (es decir, que incluyan igual cantidad de valores). De ese modo, cuando se compare la cantidad de observaciones de dos intervalos diferentes, se estarán comparando dos elementos de igual tamaño.

Al decidir la formación de los intervalos se deben considerar tres principios fundamentales. Primero, es deseable tener entre 5 y 15 intervalos. (No debe haber demasiados intervalos ya que el objetivo de la tabla de frecuencias agrupadas es simplificar la información y transmitirla a primera vista. Por otro lado, si los intervalos son muy pocos, la tabla dejaría de ser útil en cuanto a la descripción del patrón que forman las observaciones. En general, se ha observado que la cantidad adecuada sería entre 5 y 15 intervalos).

El segundo principio es que el **tamaño del intervalo** (la cantidad de valores que incluye) debe ser un número con el que resulte fácil trabajar. Se considera que 2, 3, 5, 10, ó múltiplos de 5 ó de 10, son tamaños adecuados de intervalo. También es útil que los **límites del intervalo** (los números con los que los intervalos comienzan y terminan) se establezcan de forma tal que el límite inferior de cada intervalo sea múltiplo exacto del tamaño del intervalo. Por ejemplo, supongamos que estamos utilizando un intervalo de tamaño 2, con información cuyo valor menor posible sea cercano al 0. Los límites de intervalo apropiados serían de 0 a 2, de 2 a 4, de 4 a 6, y así sucesivamente. O bien, supongamos que estamos utilizando tamaños de intervalo de 5, con información cuyo valor menor posible sea 32. En este caso, usaríamos intervalos de 30 hasta casi 35, de 35 hasta casi 40, y así sucesivamente.

Con respecto al límite superior de los intervalos, generalmente utilizamos el valor real más alto que pueda tener la variable, y que sea inmediatamente inferior al comienzo del siguiente intervalo. En otros casos, los investigadores pueden colocar como límite superior de sus intervalos un número decimal apenas menor al límite inferior del siguiente intervalo. Es decir, si los únicos valores posibles fueran números enteros, los intervalos serían de 30 a 34, 35 a 39, y así sucesivamente; aunque también podríamos utilizar intervalos de 30 a 34,9; 35 a 39,9, y así sucesivamente. (Si los números a incluir fueran de dos decimales, deberíamos utilizar intervalos de 30 a 34,99; 35 a 39,99, y así sucesivamente).

Al determinar los intervalos se debe tener en cuenta tanto la cantidad de intervalos (el principio fundamental mencionado en primer lugar) como el tamaño de los mismos (el segundo principio fundamental). A veces es difícil cumplir con ambos principios, y entonces se deben hacer concesiones. A menudo existe más de un modo de preparar una tabla de frecuencias agrupadas que cumpla con las reglas que mencionamos anteriormente (y algunas veces no es posible realizar una tabla de frecuencias agrupadas sin ajustar levemente alguna de dichas reglas.) En esos casos, tendremos la oportunidad de utilizar la creatividad, recordando siempre que el objetivo es realizar una tabla de frecuencias agrupadas que presente las observaciones de forma simple y directa.

A continuación, describimos cuatro pasos a seguir para construir una tabla de frecuencias agrupadas que cumpla con los principios que hemos estudiado:

1. Restar el valor menor al mayor para obtener la amplitud de la serie de valores. En la medición de estrés, la amplitud es 10 valores ($10 - 0 = 10$). En el ejemplo sobre la cantidad de interacciones, la amplitud es 48 ($48 - 0 = 48$).

2. Dividir la amplitud por un tamaño de intervalo razonable. (Siempre que sea posible, es conveniente utilizar los tamaños 2, 3, 5, 10, ó un múltiplo de 10). Probar distintos tamaños de intervalo hasta lograr, después de redondear, una cantidad razonable de intervalos (en general, entre 5 y 15). En el ejemplo de medición de estrés, al dividir la amplitud 10 por un tamaño de intervalo de 2, el resultado es 5, que es la menor cantidad aceptable de agrupaciones. En el ejemplo de la cantidad de interacciones sociales, al dividir la amplitud de 48 por un tamaño de intervalo de 5 y luego redondear, el resultado es 10, que es una cantidad de intervalos adecuada.

3. Realizar una lista de los intervalos de mayor a menor, controlando que el límite inferior de cada intervalo sea múltiplo exacto del tamaño del intervalo. En el caso de la medición de estrés, el primer intervalo es 10-11,9, el siguiente es 8-9,9, y así sucesivamente. Los límites inferiores de estos intervalos, 10 y 8, son múltiplos de 2, es decir, el tamaño de intervalo. (Si se utilizaran, por ejemplo, intervalos de 9-10,9 y 7-8,9, los intervalos respetarían el tamaño de 2 valores, pero los límites inferiores 9 y 7 no serían múltiplos de 2). En el caso de las interacciones sociales, cada intervalo comienza con un múltiplo de 5.

4. Proceder de igual modo que en el caso de una tabla de frecuencias no agrupadas común: leyendo una por una las observaciones, realizando una marca al lado de cada intervalo y así sucesivamente, para realizar luego una tabla prolija. (Si ya se ha realizado una tabla de frecuencias común con las observaciones, este paso resultará mucho más simple: sólo es necesario sumar las frecuencias de los valores que forman cada intervalo).

Otro ejemplo de construcción de una tabla de frecuencias agrupadas

Las tablas de frecuencias agrupadas son especialmente útiles cuando los valores incluyen decimales ya que, por lo general, cuando se trabaja con decimales, cada individuo presenta una observación diferente. Analicemos el siguiente ejemplo ficticio, basado vagamente en el trabajo de Inhoff, Lima y Carroll (1984). El objetivo de los estudios realizados consistía en observar el efecto del contexto en la velocidad de lectura de oraciones ambiguas. Las oraciones eran ambiguas, por lo cual podían ser tomadas literal o metafóricamente, como por ejemplo, "las carteleras son verrugas en el paisaje". En un capítulo posterior analizaremos este estudio más profundamente, pero por ahora nos concentraremos en la parte de la investigación que sólo pretendía determinar el tiempo de lectura de oraciones básicas ambiguas (metafóricas) dentro de un contexto no metafórico. El tiempo de lectura se mide mediante un dispositivo electrónico que, a la vez que presenta el material de lectura, determina si los ojos se están moviendo o están fijos. A los efectos del ejemplo, supondremos que los participantes en la investigación eran 100 alumnos secundarios y que cada uno debía leer varias oraciones ambiguas. Más abajo se detalla la lista (de observaciones ficticias) con el número promedio de segundos que demoró cada participante en leer las oraciones ambiguas.

2,72; 2,84; 2,63; 2,51; 2,54; 2,98; 2,61; 2,93; 2,87; 2,76; 2,58; 2,66; 2,86; 2,86; 2,58; 2,60; 2,63; 2,62; 2,73; 2,80; 2,79; 2,96; 2,58; 2,50; 2,82; 2,83; 2,90; 2,91; 2,87; 2,87; 2,74; 2,70; 2,52; 2,75; 2,99; 2,66; 2,58; 2,71; 2,51; 2,87; 2,87; 2,75; 2,85; 2,61; 2,54; 2,73; 2,96; 2,90; 2,75; 2,76; 2,93; 2,64; 2,85; 2,70; 2,56; 2,51; 2,83; 2,79; 2,76; 2,75; 2,86; 2,58; 2,87; 2,89; 2,89; 2,52; 2,59; 2,54; 2,54; 2,85; 2,83; 2,96; 2,93; 2,89; 2,92; 2,98; 2,59; 2,81; 2,78; 2,95; 2,96; 2,95; 2,56; 2,59; 2,87; 2,84; 2,84; 2,80; 2,65; 2,70; 2,61; 2,89; 2,83; 2,85; 2,52; 2,66; 2,74; 2,73; 2,88; 2,85

A continuación, se describen los pasos que se deben seguir para construir una tabla de frecuencias agrupadas utilizando los datos contenidos en la lista.

1. Restar el valor menor al mayor para saber cuál es la amplitud de la serie de valores. El valor mayor (2,99) menos el menor (2,50) da 0,49.
2. Dividir la amplitud por varios tamaños de intervalo posibles hasta encontrar, después de redondear, una cantidad razonable de intervalos. Cuando la amplitud es pequeña, es necesario tener en cuenta tamaños de intervalo representados por decimales. Sin embargo, aun utilizando decimales, es conveniente utilizar sólo tamaños de intervalos que sean números comunes y regulares. Así, en este ejemplo, podríamos tomar un tamaño de intervalo de 0,1, lo cual daría como resultado 5 intervalos, pero sería aún más adecuado utilizar el tamaño 0,05, para obtener 10 intervalos.
3. Realizar una lista de los intervalos ordenándolos de mayor a menor. En este caso, los intervalos deberían comenzar con 2,95-2,99 y continuar hasta 2,50-2,54.
4. Proceder del mismo modo que con una tabla de frecuencias común. La tabla 1-5 muestra el resultado.

Tabla 1-5.
Tabla de frecuencias agrupadas correspondiente a un estudio realizado a 100 alumnos secundarios, para medir el tiempo promedio de lectura de oraciones ambiguas (en segundos).

Tiempo de lectura	Frecuencia
2,95–2,99	9
2,90–2,94	7
2,85–2,89	20
2,80–2,84	11
2,75–2,79	10
2,70–2,74	10
2,65–2,69	4
2,60–2,64	8
2,55–2,59	10
2,50–2,54	11

HISTOGRAMAS

Los gráficos constituyen otro medio adecuado para facilitar la comprensión de una cantidad importante de registros. "Una imagen vale más que mil palabras", y a veces más que mil números. Una manera de graficar la información de una tabla de frecuencias es con un gráfico de barras especial denominado **histograma**. En un histograma, la altura de cada barra representa la frecuencia que le corresponde al intervalo de acuerdo con la tabla de frecuencias. Además, las barras están ubicadas una al lado de la otra, sin espacios entre ellas. Los histogramas se parecen al contorno de una ciudad en el horizonte. La figura 1-3 muestra dos histogramas basados en el ejemplo de medición del estrés (uno representa la tabla de frecuencias clásica y el otro la tabla de frecuencias agrupadas).

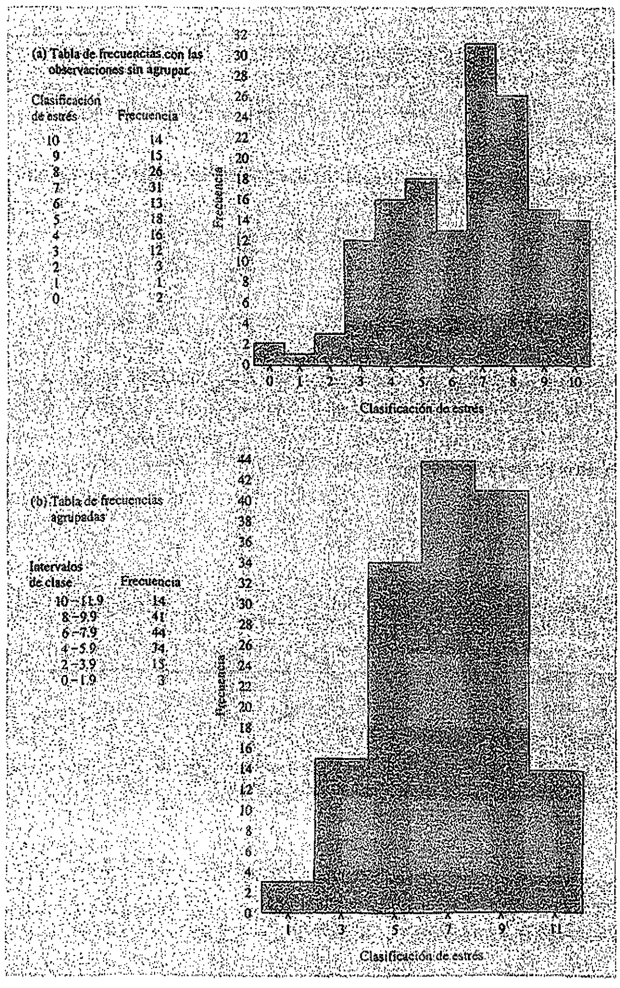


Figura 1-3. Histograma basado en (a) tabla de frecuencias y (b) tabla de frecuencias agrupadas según información de Aron, Paris, & Aron (1995).

Cuadro1-2.
La angustia matemática, la angustia estadística y el alumno: un mensaje para aquellos que están realmente preocupados por este curso.

"Afrontémoslo: muchos alumnos le temen a este curso, incluso hasta el punto de convertirse en un caso de "angustia estadística". (Zeidner, 1991) Por lo tanto, si hay quienes se pondrán tensos a medida que comiencen a aparecer los números, será mejor tratar el tema ahora mismo.

En primer lugar, este curso es una oportunidad para comenzar de cero con los dígitos. El desempeño anterior (o la ausencia de desempeño) en geometría, trigonometría, cálculo u horrores semejantes no debe influir de ningún modo en la forma de encarar la estadística. Es una materia muy distinta de las otras.

En segundo lugar, si la preocupación persiste, es necesario descubrir cuál es su origen. La angustia matemática o estadística, la angustia provocada por los exámenes, la angustia en general y el bajo nivel de confianza en uno mismo son elementos que aparentemente influyen en el nivel de dificultad que enfrentan los alumnos en los cursos de matemática. (Cooper & Robinson, 1989; Swinell & Higbee, 1991)

Angustia matemática: si el problema es la angustia matemática, recomendamos averiguar si el centro de asesoramiento de la facultad cuenta con algún programa de ayuda para aquellos que tienen esta dificultad (como por ejemplo, los programas estudiados por Schneider & Nevid, 1993; Vance & Watson, 1994). De lo contrario, existe un buen libro sobre el tema: *¿Cómo tener éxito con las matemáticas?: guía para que cada alumno pueda superar la angustia matemática [Succeed With Math: Every Student's Guide to Conquering Math Anxiety].* (1987)

Tobías, una alumna que solía evitar la matemática, sugiere que el objetivo debe ser la "salud mental matemática", a la que ella misma define como "la voluntad de aprender las matemáticas que necesitas, cuando las necesitas". (p.12) (Tal vez este curso de estadística pueda ser una de esas oportunidades).

Tobías explica que la salud mental matemática se pierde generalmente en la primaria. Cuando el alumno pasa al frente, su mente se bloquea y le resulta imposible encontrar la respuesta correcta a un problema aritmético. Si luego de esa experiencia el alumno conserva alguna confianza, probablemente la pierda durante los exámenes cronometrados, los cuales, aunque no resulte evidente, son difíciles para todos, excepto para unos pocos expertos.

Tobías sostiene que los alumnos hábiles para las matemáticas no son necesariamente más inteligentes que el resto, sino que realmente conocen cuáles son sus fortalezas y sus debilidades, sus estilos de pensamientos y sentimientos con respecto a un problema. No se juzgan severamente por sus errores. Particularmente, no esperan comprender las cosas en forma instantánea. Permitirse tener un "aprendizaje lento" no significa ser menos inteligente, implica que la propia salud mental matemática está creciendo.

Uno de los trucos de Tobías es dividir la página en dos y trabajar con los problemas de estadística en la mitad derecha. Cuando la angustia estadística bloquea su trabajo, utiliza la mitad izquierda para anotar sus pensamientos. En general, se trata de pensamientos negativos, como por ejemplo "nunca voy a aprender esto". Luego, Tobías reemplaza esos pensamientos por otros más razonables y positivos, tales como "sólo es cuestión de tiempo", y "si me trabo durante demasiado tiempo siempre habrá alguien que me ayude" o "mira todo lo que ya he aprendido sobre estadística". De ese modo, nunca deja de trabajar (ya sea en el problema o en los obstáculos psicológicos que le impiden resolverlo).

Angustia por los exámenes: si el problema es la angustia causada por los exámenes, también podemos brindar nuestra ayuda. Es posible que en el centro de asesoramiento dicten algún curso sobre el tema o exista algún libro al respecto. Sin embargo, creemos que resultaría útil para el alumno saber que existen tres modos de reducir la angustia causada por los exámenes, de tal forma que pueda elegir la combinación de los tres que más se adecue a cada uno.

El primer método parte del supuesto de que dominar un tema cualquiera pesa más que la angustia que pueda sentirse; incluso existen investigaciones que así lo sugieren (Klejin, van der Ploeg. & Topman, 1994).

Además, el hecho de estar bien preparado tiene que reducir la angustia. El mejor momento para comenzar a aplicar esta táctica es el primer examen del curso: no habrá material antiguo para repasar, el éxito no dependerá de haber comprendido temas previos y será útil para un mejor desempeño a lo largo de todo el curso. (Hasta se podría conquistar la simpatía del profesor o ayudante de cátedra presentándole una lista de los temas estudiados, explicándole por qué se está siendo tan exigente y averiguando si falta estudiar algún tema). La preparación para el examen debe ser absurdamente completa, pero sólo para unos pocos exámenes. Después de tener éxito en ellos, la angustia debería disminuir.

El segundo método para disminuir la angustia provocada por los exámenes supone que el problema es provocado por la ansiedad física y por un estado emocional general de angustia. Cualquier tipo de angustia produce ansiedad, y una de las relaciones más comprobadas en psicología es la que existe entre la ansiedad y el desempeño (*véase* figura 3-6). Mientras que la ansiedad moderada es positiva para el desempeño, demasiada (o muy poca) ansiedad lo reduce drásticamente.

Cuando la raíz del problema es la ansiedad, es útil para el alumno saber que no existe nada malo en el "hardware", nada malo ocurre con su cerebro, su inteligencia o su forma de estudiar el material. El alumno debe confiar en que realmente conoce el tema. Mueller, Elser y Rollack (1993) lo demostraron probando directa e indirectamente en qué medida era adecuado el aprendizaje de una lista de palabras, descubriendo que al ser probados en forma directa, los sujetos que estaban angustiados mostraban un desempeño inferior al de aquellos que no lo estaban. Sin embargo, al ser probados indirectamente, el nivel de desempeño de unos y otros era prácticamente el mismo.

¿Cómo se puede reducir la angustia y la ansiedad? Una vez más, el centro de asesoramiento debería estar capacitado para brindar ayuda en la resolución de este problema, o bien, aconsejar la lectura de libros útiles al respecto. Existen muchas técnicas, tales como aprender a respirar de forma adecuada y tomar un breve descanso para relajarse profundamente. Después de aprender el método de relajación se puede intentar la "desensibilización sistemática", a través de la cual el alumno se imagina a sí mismo en situaciones de examen cada vez más complejas mientras permanece físicamente relajado. Lo importante es guiarse por lo que resulte más adecuado para cada uno.

Un truco muy útil para reducir la ansiedad es crear una práctica de examen, tan parecida a una prueba real como sea posible, de tal manera de familiarizarse con la mayor cantidad de aspectos de un examen (de modo que causen menos ansiedad). Es importante que el alumno se esfuerce por reproducir los aspectos que más lo inquieten. Si el problema es la falta de tiempo, una vez que se considere bien preparado debe establecer un límite de tiempo para resolver algunos de los ejercicios que tenga como tarea. Las respuestas deben ser completas y legibles, ya que estos dos requisitos pueden ser los que lo hagan sentir lento durante un examen. Si lo que al alumno le molesta es la presencia de otras personas o el sonido de los lápices avanzando rápidamente mientras el suyo se encuentra suspendido en el aire, entonces debe realizar la práctica de examen con otros alumnos de su curso. Incluso es recomendable proponer explícitamente una competencia para comprobar quién termina primero.

Finalmente, si el profesor puede realizar los arreglos necesarios, probablemente el desempeño del alumno mejore si realiza exámenes no cronometrados. En este tipo de exámenes, los alumnos que sufren angustia obtienen los mismos resultados que los demás; mientras que en exámenes cronometrados, obtienen peores resultados que el resto (Onwuegbuzie, 1994).

El tercer método para enfrentar la angustia producida por los exámenes es cognoscitivo. Éste parte del supuesto de que se necesita reemplazar cualquier pensamiento negativo irracional, que surja al estudiar o dar un examen, por pensamientos positivos y racionales. Una vez más, el centro de asesoramiento debería estar capacitado para ayudar al respecto. Si no es así, el método de Tobías también es adecuado en este caso.

Poca confianza en uno mismo: si el alumno sospecha que su problema es una falta general de confianza, o si existe alguna otra cosa en su vida que lo está preocupando, nuevamente sugerimos que es momento de intentar obtener ayuda del cordial centro de asesoramiento universitario.

Un último comentario sobre la angustia y la ansiedad: es probable que el lector se encuentre dentro del 15 ó 20% de seres humanos (y animales superiores) que nacen con una cualidad que los hace más propensos a percibir las estimulaciones sutiles, lo cual con frecuencia los hace sumamente intuitivos, e incluso dotados. No obstante, estos seres también resultan fácilmente excitables por niveles de estimulación que no perturban a otros. (Eysenck, 1981; Kagan, 1994). Nuestra propia investigación (E. Aron, 1996; E. Aron & A. Aron, 1997) sobre individuos con estas cualidades especiales, a quienes llamamos PAS **(personas altamente sensibles)**, indica que las mismas no son ni particularmente neuróticas ni excesivamente emocionales. En verdad, poseen muchas cualidades positivas: en comparación con otros, muestran mayores niveles de equidad, una mayor apreciación del arte y la música y una vida interior rica y compleja. Se trata de un conjunto: si una persona posee mayor sensibilidad, también es más fácilmente excitable.

Si el lector considera que es una PAS, tal vez ayude a disminuir su ansiedad el hecho de comprender por qué es más excitable que los demás en las mismas situaciones. ¡Al menos no es necesario preocuparse por estar preocupado! Muchas PAS no tienen problemas con los exámenes, pero usualmente existe algún aspecto de su vida en el que la ansiedad les cause problemas (deportes, hablar en público, etc.); puede ser que en el caso del lector el aspecto conflictivo sean los exámenes. Cabe recordar que lo importante es lo que uno realmente sabe, que probablemente sea bastante.

Así que buena suerte a todos. Les deseamos lo mejor durante este curso y en todos los aspectos de sus vidas.

¿Cómo confeccionar un histograma?

A continuación presentamos los cuatro pasos para la construcción de un histograma:

1. Construir una tabla de frecuencias (o una tabla de frecuencias agrupadas).

2. Ubicar la escala de intervalos al pie de la página. Los números deben ir de izquierda a derecha y de menor a mayor. En el caso de una tabla de frecuencias agrupadas, el histograma será de intervalos. Comúnmente, en un histograma basado en una tabla de frecuencias agrupadas, sólo se marca el punto medio de cada intervalo en el centro de la base de cada barra. El punto medio es el centro del intervalo, a mitad del camino entre el comienzo de un intervalo y el siguiente. (Para obtener el punto medio se debe restar el límite inferior del intervalo en cuestión al límite inferior del intervalo siguiente; dividir el resultado por 2 y sumarlo al límite inferior del intervalo, cuyo punto medio estamos determinando).

3. Desplegar una escala de frecuencias a lo largo del margen izquierdo de la página. La escala debe partir de 0, en la parte inferior, hasta la mayor de todas las frecuencias de los intervalos.

4. Dibujar una barra para cada intervalo. La altura de cada barra es la frecuencia del intervalo sobre la cual se ubica esa barra.

Resulta más fácil construir un histograma utilizando papel para gráficos.

Otros ejemplos de histogramas

La figura 1-4 muestra un histograma construido a partir de la tabla de frecuencias agrupadas correspondiente al ejemplo de la cantidad de interacciones sociales vividas por estudiantes universitarios durante una semana. La figura 1-5 muestra el histograma basado en la tabla de frecuencias agrupadas correspondiente al ejemplo referido al tiempo de lectura de una oración ambigua.

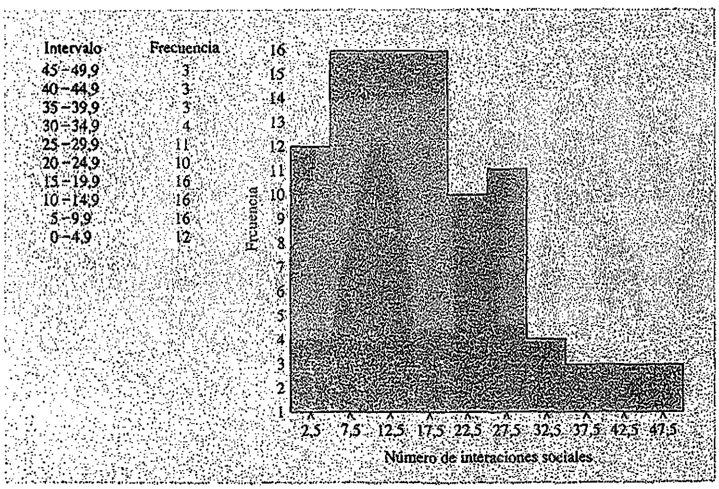


Figura 1-4. Histograma que representa el número de interacciones sociales vividas durante una semana por 94 estudiantes universitarios, basado en frecuencias agrupadas. (**Fuente**: McLaughlin-Volpe et al., 1998).

Tiempo de lectura	Frecuencia
2,95 – 2,99	9
2,90 – 2,94	7
2,85 – 2,89	20
2,80 – 2,84	11
2,75 – 2,79	10
2,70 – 2,74	10
2,65 – 2,69	4
2,60 – 2,64	8
2,55 – 2,59	10
2,50 – 2,54	11

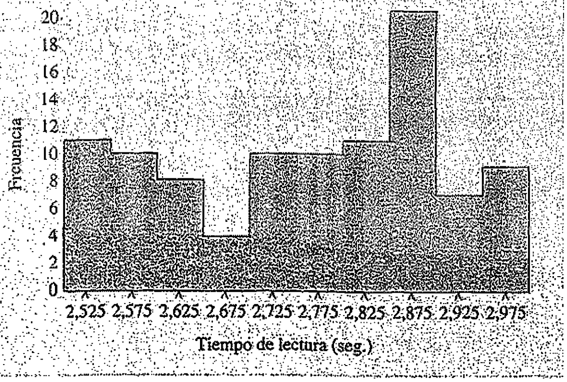


Figura 1-5. Histograma que representa el tiempo promedio de lectura de oraciones ambiguas por parte de 100 alumnos secundarios, basados en frecuencias agrupadas (datos ficticios).

POLÍGONOS DE FRECUENCIAS

Existe otro método utilizado comúnmente para mostrar gráficamente la información contenida en una tabla de frecuencias. Este tipo de gráfico, denominado **polígono de frecuencias**, es básicamente la versión del histograma representado con un gráfico de líneas. En lugar de barras, la frecuencia de cada intervalo se indica a través de la altura de una línea que se desliza por la página, creando una especie de contorno de montañas. La figura 1-6 muestra los polígonos de frecuencias creados a partir de las tablas de frecuencias comunes y agrupadas correspondientes al ejemplo de medición del estrés.

¿Cómo confeccionar un polígono de frecuencias?

A continuación describimos los cinco pasos que se deben seguir para la creación de un polígono de frecuencias:

1. Realizar una tabla de frecuencias (o una tabla de frecuencias agrupadas).

2. Ubicar la escala de intervalos al pie de la página. Incluir un intervalo extra al principio y otro al final de la serie de intervalos que tienen observaciones realmente obtenidas. Los intervalos adicionales son necesarios para asegurar que la línea comience y termine en la base del gráfico, es decir, en la frecuencia cero. Al incluir los intervalos adicionales se crea una figura cerrada o "polígono". Al igual que con los histogramas, la escala se construye marcando sólo el punto medio de cada intervalo.

3. Realizar una escala de frecuencias a lo largo del margen izquierdo de la página. La escala debe partir de 0, en la parte inferior, hasta la mayor de las frecuencias de todos los intervalos.

4. Marcar un punto sobre el centro de cada intervalo, a la altura correspondiente a la frecuencia de ese intervalo.

5. Unir los puntos por medio de líneas.

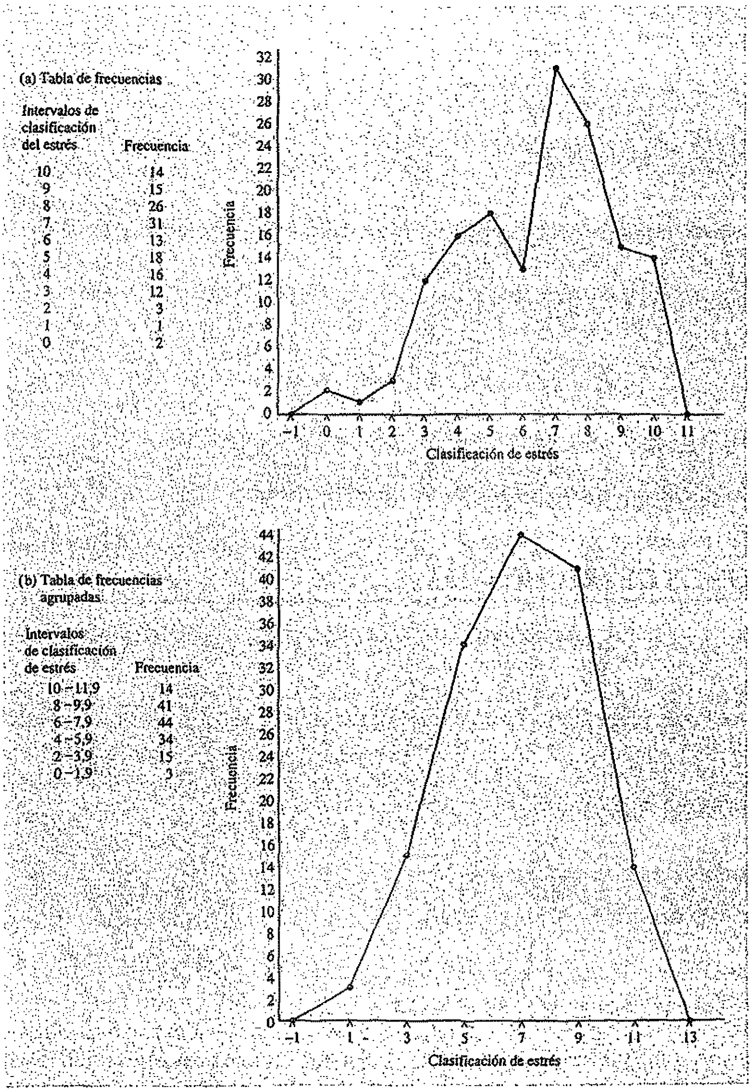


Figura 1-6. Polígonos de frecuencias basados en (a) una tabla de frecuencias y (b) una tabla de frecuencias agrupadas, construidas con los datos de Aron, Paris & Aron (1995).

Otro ejemplo de polígonos de frecuencias

La figura 1-7 muestra los cinco pasos necesarios para construir un polígono de frecuencias, utilizando la tabla de frecuencias agrupadas correspondiente al ejemplo sobre interacciones sociales de varios alumnos.

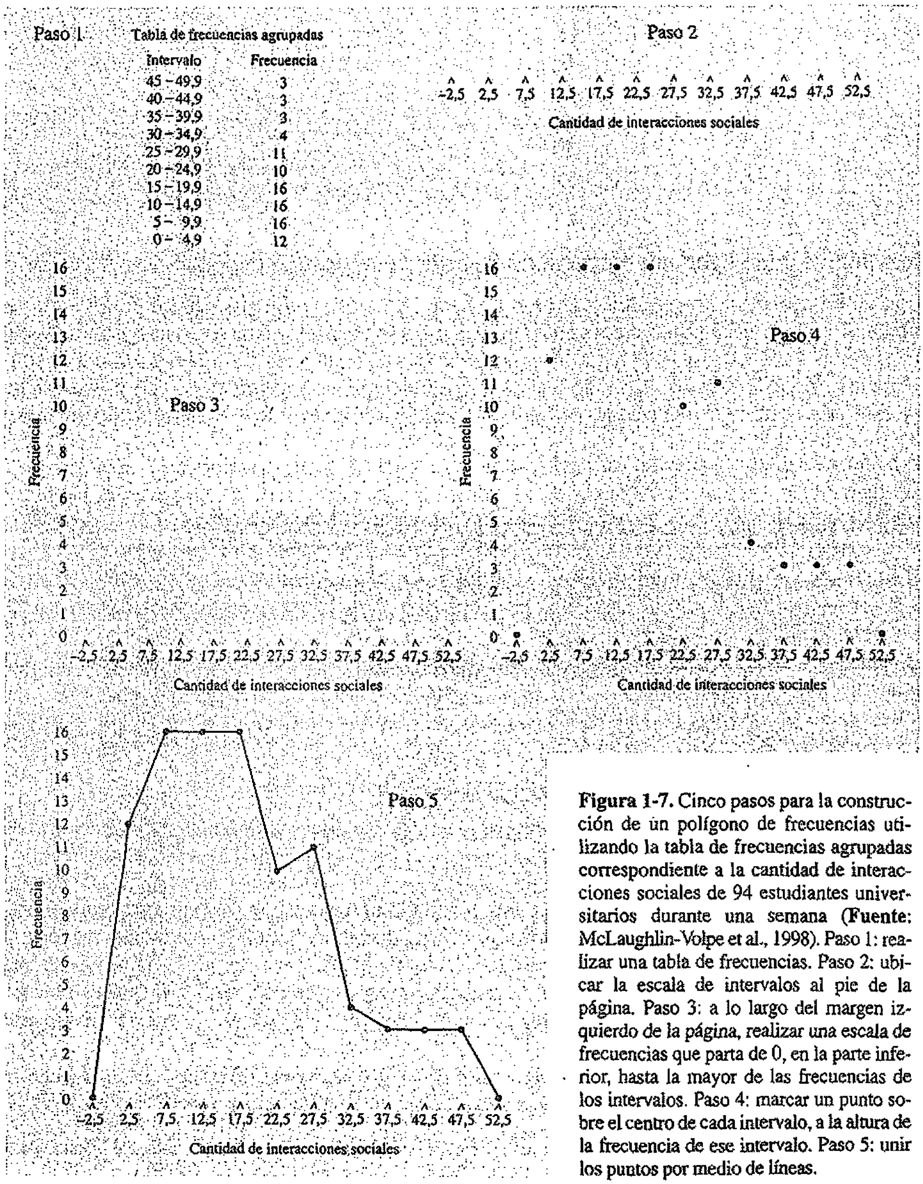
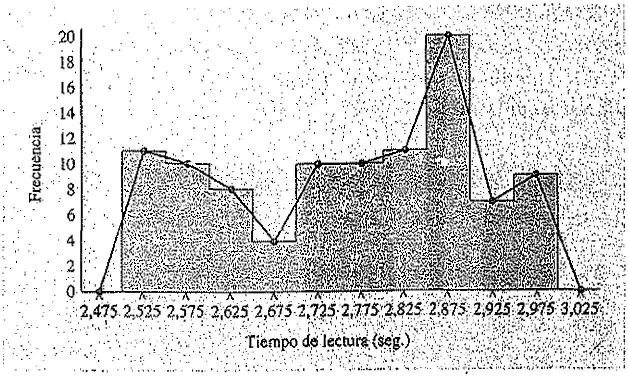


Figura 1-7. Cinco pasos para la construcción de un polígono de frecuencias utilizando la tabla de frecuencias agrupadas correspondiente a la cantidad de interacciones sociales de 94 estudiantes universitarios durante una semana (**Fuente:** McLaughlin-Volpe et al., 1998). Paso 1: realizar una tabla de frecuencias. Paso 2: ubicar la escala de intervalos al pie de la página. Paso 3: a lo largo del margen izquierdo de la página, realizar una escala de frecuencias que parta de 0, en la parte inferior, hasta la mayor de las frecuencias de los intervalos. Paso 4: marcar un punto sobre el centro de cada intervalo, a la altura de la frecuencia de ese intervalo. Paso 5: unir los puntos por medio de líneas.

Figura 1-8. Creación de un polígono de frecuencias a partir de un histograma, utilizando el histograma correspondiente al tiempo promedio de lectura de oraciones ambiguas por parte de 100 alumnos secundarios (datos ficticios).



Creación de un polígono de frecuencias a partir de un histograma

Si ya se ha hecho un histograma, todo lo que hace falta para convertirlo en un polígono de frecuencias es colocar un punto en el centro de la parte superior de cada barra y luego unirlos entre sí. La única complicación es agregar los puntos medios de cada intervalo vacío al principio y al final de la serie de intervalos, de modo que el polígono comience y termine en 0. La figura 1-8 ilustra este procedimiento utilizando el ejemplo basado en el estudio del tiempo de lectura de oraciones ambiguas.

FORMAS DE LAS DISTRIBUCIONES DE FRECUENCIAS

Una tabla de frecuencias, un histograma o un polígono de frecuencias describen una **distribución de frecuencias**. Es decir, muestran el patrón conforme al cual las frecuencias se dispersan o "distribuyen".

A los psicólogos también les resulta útil describir con palabras el patrón de distribución de las frecuencias. En general, las descripciones se refieren a aspectos de la forma del histograma o del polígono de frecuencias. En esta sección analizamos estos aspectos y su terminología específica.

Distribuciones de frecuencias unimodales y bimodales

Un aspecto importante relacionado con la forma de una distribución de frecuencias es el hecho de que la figura presente un solo punto máximo principal (una "torre" alta en el histograma o un "pico" alto principal en el polígono de frecuencias). En el ejemplo de las clasificaciones de estrés, el registro más frecuente es el 7, lo que da como resultado un gráfico con sólo un área muy elevada. Este tipo de distribución se denomina **unimodal**. Una distribución con dos puntos elevados prácticamente iguales es una distribución **bimodal**. Cualquier distribución con dos o más puntos elevados se denomina multimodal.[2] Finalmente, una distribución en la que todos los valores presentan prácticamente la misma frecuencia se denomina **rectangular**. La figura 1-9 muestra las distintas distribuciones de frecuencias.

En general, la información que recolectamos en investigaciones psicológicas es prácticamente unimodal. Las distribuciones bimodales y multimodales se dan sólo ocasionalmente.

[2] Estrictamente hablando, una distribución es bimodal o multimodal sólo si los picos son exactamente iguales. Pero es común en la práctica utilizar estos términos más informalmente para describir la forma general.

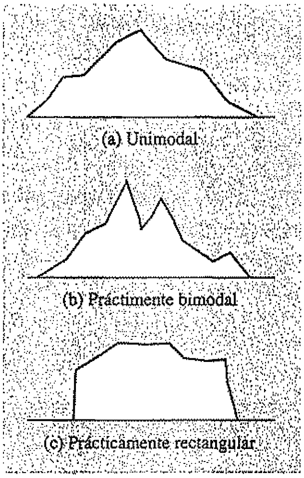


Figura 1-9. Ejemplos de polígonos de frecuencias (a) unimodal, (b) prácticamente bimodal y (c) prácticamente rectangular.

Un ejemplo bimodal sería la distribución de la cantidad de empleados en cuyos gerentes de nivel superior se han fijado en ellos por algún motivo. Si construyéramos una distribución de frecuencias con respecto a la calidad del trabajo de dichos empleados, los puntos altos en un gráfico de este tipo coincidirían con los valores que indican una calidad de trabajo muy pobre o muy buena. Un ejemplo de distribución rectangular sería la cantidad de niños en cada grado de la escuela primaria. En este caso, habría prácticamente la misma cantidad en primer grado que en segundo, y así sucesivamente. La figura 1-10 ilustra estos ejemplos.

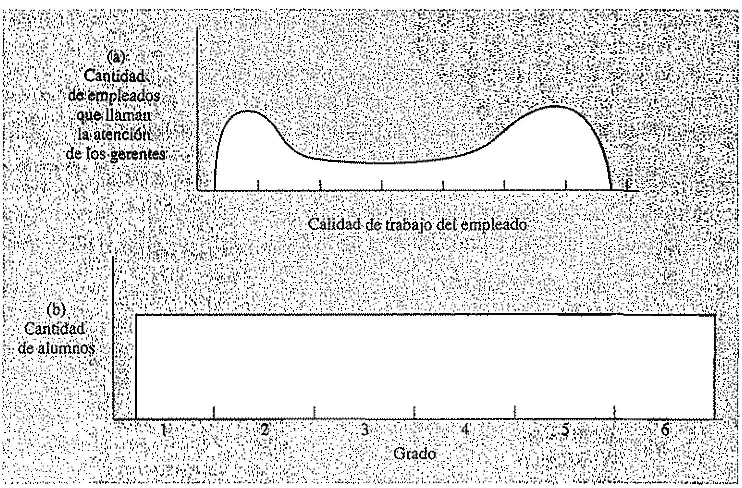


Figura 1-10. Ejemplos ficticios de distribuciones que no son unimodales. (a) Distribución bimodal que indica las posibles frecuencias en diferentes niveles de calidad del trabajo realizado por empleados que llamaron la atención de gerentes de mayor nivel. (b) Distribución rectangular que muestra las posibles frecuencias de la cantidad de alumnos en los diferentes grados de la escuela primaria.

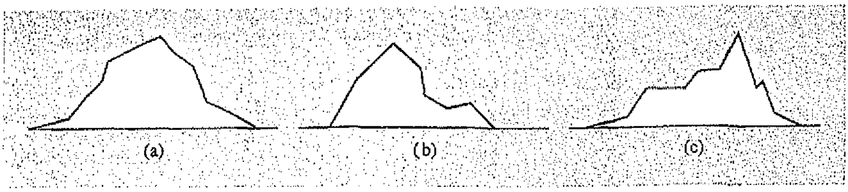


Figura 1-11. Ejemplos de polígonos de frecuencias de distribuciones (a) prácticamente simétricas, (b) asimétricas hacia la derecha (positivamente asimétricas) y (c) asimétricas hacia la izquierda (negativamente asimétricas).

Distribuciones simétricas y asimétricas

Otro aspecto para observar en el ejemplo de puntuaciones de estrés es que la distribución era ladeada, con mayor cantidad de casos cerca del extremo correspondiente al valor más alto. Esto es algo poco frecuente: la mayoría de los fenómenos que medimos en psicología tienden a presentar prácticamente la misma cantidad de casos a ambos lados del centro. Es decir, la mayoría de las distribuciones son prácticamente **simétricas** (si las dobláramos por la mitad, las dos mitades serían iguales).

Las distribuciones que claramente no son simétricas se denominan **asimétricas**. La distribución de las puntuaciones de estrés es un ejemplo de distribución asimétrica, es decir, presenta un lado largo y estirado, como una especie de cola. El lado con **menor cantidad** de casos (el lado que parece una cola) es el lado al que nos referimos para nombrar la dirección de la asimetría. Una distribución como la del ejemplo de medición del estrés, que presenta muy pocos casos en el extremo correspondiente a los valores bajos, es una distribución asimétrica hacia la izquierda. El ejemplo relacionado con el tiempo de lectura también es asimétrico hacia la izquierda. Por otro lado, el ejemplo relacionado con las interacciones sociales es asimétrico hacia la derecha. La figura 1-11 ilustra casos de distribuciones simétricas y asimétricas.

Una distribución asimétrica hacia la derecha se denomina también **positivamente asimétrica**. Una distribución asimétrica hacia la izquierda se denomina también **negativamente asimétrica.**

En la práctica, las distribuciones muy asimétricas aparecen en las investigaciones psicológicas, principalmente cuando lo que se está midiendo presenta un límite superior o inferior. Por ejemplo, en los Estados Unidos, la distribución de la cantidad de hijos por familia es asimétrica

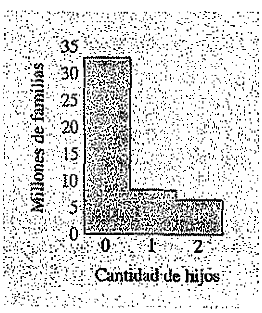


Figura 1-12. Distribución asimétrica hacia la derecha: cantidad de hijos en los EE.UU. por cada familia, en 1988. (**Fuente:** Oficina de Censos de EE.UU. [U.S. Bureau of the Census], 1990).

hacia la derecha (*véase* figura 1-12), porque no es posible tener menos de cero hijos. El **efecto piso** es la acumulación de observaciones en el extremo izquierdo del gráfico, debido a que es imposible que exista un registro menor. El ejemplo referido a la cantidad de interacciones sociales también presenta un efecto piso, porque nadie puede tener menos de 0 interacciones.

La figura 1-13 ilustra un caso de distribución asimétrica causada por la existencia de un límite superior. La distribución, referida al resultado logrado por adultos en una prueba sobre las tablas de multiplicar, es drásticamente asimétrica hacia la izquierda. La mayoría de los resultados se acumularon a la derecha, en el extremo con los valores más altos (resultado perfecto.) El ejemplo descripto presenta un **efecto techo**. El ejemplo de medición del estrés también presenta un leve efecto techo. Esto se debe a que muchos alumnos sufrían de un alto nivel de estrés –la puntuación máxima era 10– pero, por lo general, a las personas no les gusta utilizar las máximas.

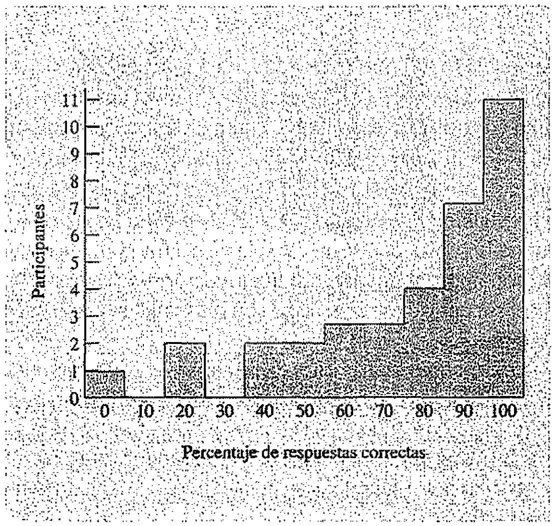


Figura 1-13. Distribución asimétrica hacia la izquierda: distribución ficticia de los resultados obtenidos por adultos en una prueba sobre las tablas de multiplicar.

Distribuciones normales y cúrticas

Finalmente, una distribución puede describirse según lo "espeso" o "pesado" de sus colas (cuán altas son las colas, cuántos casos están incluidos en ellas). El patrón de comparación es una curva con forma de campana, a la que se aproximan la mayoría de las distribuciones de frecuencias provenientes de investigaciones psicológicas y de la naturaleza en general, que se denomina **curva normal**, y a la que dedicaremos considerable atención en capítulos posteriores. Sin embargo, por ahora es suficiente destacar que la curva normal es unimodal y simétrica, lo cual tiene el tipo de forma de campana que presenta la figura 1-14a. Los tres ejemplos principales de este capítulo se aproximan a una curva normal en un sentido muy general, aunque, como hemos visto, todos ellos son algo asimétricos. Según nuestra experiencia, la mayoría de las distribuciones que resultan de las investigaciones psicológicas realmente se aproximan más a la curva normal que estos tres ejemplos.

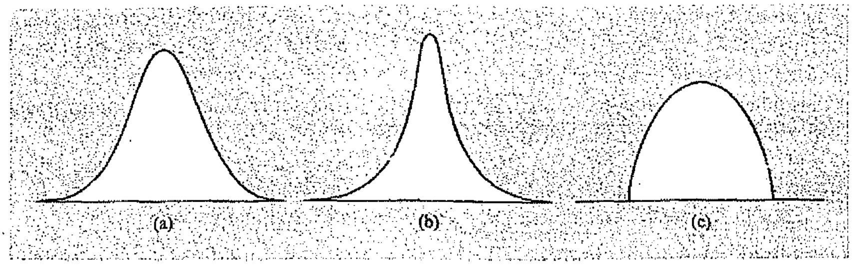


Figura 1-14. Ejemplos de distribuciones (a) normal, (b) de colas espesas y (c) de colas delgadas. (Adaptación, DeCarlo 1997).

El término **curtosis** se refiere al grado en el que la forma de una distribución difiere de la curva normal, principalmente con respecto al hecho de que las colas sean más espesas o delgadas que las de la curva normal (DeCarlo, 1997). El término curtosis proviene de la palabra griega *kyrtos*, que significa "curva". La línea oscura de la figura 1-14b indica una distribución cúrtica con colas más espesas que las de la curva normal. La figura 1-14c presenta un ejemplo extremo de distribución cúrtica, una distribución sin colas. (Una distribución rectangular sería un caso aún más extremo).

Además de la diferencia en el espesor de las colas con respecto a la curva normal, las distribuciones con colas espesas por lo general son más empinadas que la curva normal (*véase* figura 1-14b), y aquéllas con colas más delgadas o sin colas, por lo general son más chatas que la curva normal (*véase* figura 1-14c).

Las distribuciones con colas espesas se ven como si a la curva normal se la pellizcara en la mitad, y parte de ella se elevara formando un pico agudo y el resto se extendiera para formar espesas colas. Las distribuciones con colas delgadas (o sin colas), se ven como si se tirara hacia fuera el centro de la distribución y se absorbieran las colas y el pico. De todos modos, aun cuando usualmente la elevación o chatura de una distribución esté relacionada con la curtosis, lo más importante es el espesor de las colas.

CONTROVERSIAS Y LIMITACIONES

La controversia más importante con respecto a la utilización de tablas de frecuencias, histogramas y polígonos de frecuencias no se genera entre los psicólogos, sino entre el público en general. La utilización y el uso incorrecto de estos procedimientos descriptivos por parte de los medios parece haber creado escepticismo con respecto a la confiabilidad de la estadística en general y de las tablas y cuadros estadísticos en particular ¡Quién no ha escuchado decir que "la estadística miente"! En realidad, las personas pueden mentir a través de la estadística, y así lo hacen. Es tan sencillo como mentir con palabras, pero las mentiras dichas con números son seguramente más difíciles de reconocer. En esta sección destacamos dos maneras a través de las cuales las tablas de frecuencias y los gráficos equivalentes pueden ser usados en forma errónea, y mostramos cómo reconocer esos usos incorrectos. (Gran parte del material está basado en la excelente y entretenida exposición de estos temas). (Tufte, 1983)

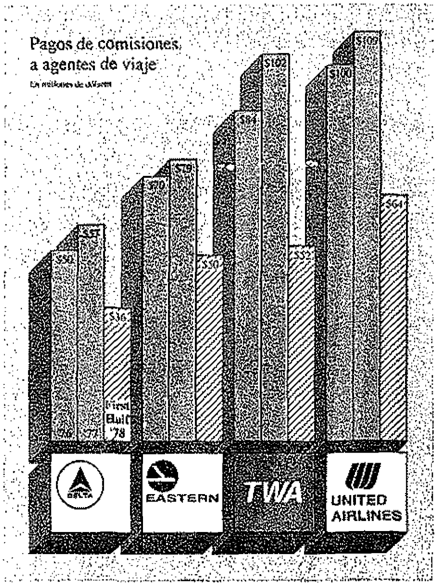


Figura 1-15. Representación engañosa de una distribución de frecuencias a causa de los distintos tamaños de intervalos. (**Fuente:** *New York Times*, 8 de agosto, 1978, p. D-1. © 1978 por la *New York Times Company*. Reimpreso con autorización).

No utilizar intervalos del mismo tamaño

Como ya observamos, un requisito indispensable de las tablas de frecuencias es que los tamaños de los intervalos sean iguales. Si los intervalos no son iguales, la tabla o gráfico resultante puede ser muy engañosa. Tufte (1983) nos da un ejemplo, ilustrado en la figura 1-15, tomado del respetable (y usualmente preciso) *New York Times*. Aparentemente, este gráfico señala que las comisiones pagadas a agentes de viaje cayeron drásticamente en el año 1978. Sin embargo, una lectura más detallada revela que la tercera barra de cada caso representa sólo el primer semestre del año 1978. Por lo tanto, se está comparando sólo la mitad de un año con cada uno de los años anteriores completos. Presumiendo que el segundo semestre del año 1978 haya sido igual al primero, la información en este gráfico sugiere en realidad que en el año 1978 se produjo un aumento y no una disminución. (Por ejemplo, las cifras estimadas por Delta Airlines para todo el año 1978 serían de $72 millones, cifra mucho mayor a los $57 millones de 1977).

Exageración de las proporciones

Comúnmente, la altura de un histograma o de un polígono de frecuencias debería comenzar en 0, como menor valor de la escala, y continuar hasta su valor mayor. Al mismo tiempo, el gráfico debería tener un ancho igual a 1,5 veces su altura. El ejemplo sobre puntuaciones de estrés ilustra este principio. Sin embargo, observemos qué sucede si hacemos el gráfico mucho más alto o corto (como lo ilustra la figura 1-16): la impresión lograda puede ser una mayor o menor diferencia entre los intervalos. El efecto es semejante al de los espejos de los parques de diversiones, pues la verdadera imagen está distorsionada.

Por supuesto que cualquier forma particular es precisa en algún sentido. Pero se ha adoptado la escala 1,5:1 para brindar un patrón de comparación. Cambiar esta proporción sería engañoso para la vista.

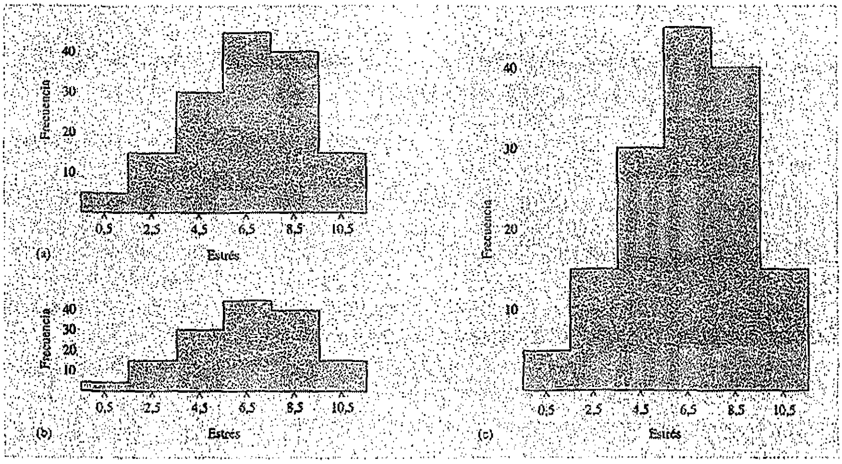


Figura 1-16. Histogramas que distorsionan las clasificaciones de estrés de los alumnos al no cumplir con la norma de uso que establece un ancho igual a 1,5 veces la altura. (Fuente: Aron, Paris & Aron).

Cuadro1-3. Sexo, etnia y desempeño matemático.

De tanto en tanto, alguien intenta argumentar que al ser los hombres blancos quienes tienden a lograr los mayores éxitos en matemática, las mujeres y las personas de color son inherentemente menos capaces en ese campo (y en la estadística). Planteamos este tema porque seguramente el alumno ha oído hablar de tales actitudes y no queremos que ellas influyan sutilmente en ninguno de ellos, especialmente en los que no son hombres blancos.

Con respecto a los sexos, existía cierta brecha entre ellos que sugería que las mujeres eran menos capaces para la matemática, pero la misma se ha estrechado sustancialmente, y aún más en aquellos lugares en los que se han mejorado la actitud y las oportunidades. Esto sugiere que las diferencias no eran genéticas sino que estaban determinadas culturalmente. (Baker & Jones, 1993; Hyde, 1993) En una revisión de las investigaciones realizadas sobre el tema, Hyde (1993) descubrió que en la secundaria los varones aun se desempeñan mejor que las mujeres en la solución de problemas complejos y en el SAT (*Scholastic Amplitude Test*, Examen de aptitud escolar) por diversas razones (es como el hecho de que los varones continúen tomando cursos de matemática). Sin embargo, estas diferencias no se encontraban en otros indicadores de capacidad matemática. Algunos han descubierto que los varones tienen mejor razonamiento espacial, pero este aspecto aún no ha sido comprendido o medido adecuadamente, y es probable que no se deba a una diferencia genética.

Es verdad que, en general, los que mejor se han desempeñado en el campo de la matemática han sido hombres. Pero aun así, las diferencias son leves, y las mujeres no son las más propensas a tener los peores desempeños, como ocurriría si existieran diferencias genéticas. Más que nunca, las mujeres están obteniendo doctorados en matemática, aunque también es el campo con el mayor índice de deserción femenina. Tal vez una de las razones sea que, si bien las mujeres no consideran la matemática un campo "para hombres", los hombres definitivamente sí lo consideran de ese modo. (Hyde et al., 1990) Con esa actitud, es posible que los hombres dedicados a la matemática no incentiven a sus alumnas o colegas femeninas. No obstante, cambiar ese tipo de actitud sólo es cuestión de tiempo.

En lo que respecta al desempeño femenino en estadística, chequeamos las calificaciones en nuestras propias clases de introducción y realmente no encontramos ninguna diferencia confiable relacionada con el sexo. Tampoco Buck (1985) encontró dicha diferencia en un análisis de trece semestres de cursos de estadística para alumnos universitarios no graduados, principiantes y avanzados.

En cuanto a los grupos étnicos, sí existen diferencias de desempeño, pero ninguna que no pueda ser explicada a través de las diferentes oportunidades de cada grupo. En particular, la gente de color no es alentada con frecuencia a estudiar matemática de alto nivel. Y lo que es peor, las escuelas a las que asisten por lo general cuentan con menos cursos avanzados de matemática, menos profesores calificados en la materia y menos recursos para la enseñanza de la matemática y la ciencia. La falta de estudiantes doctorados dentro de estas comunidades probablemente perpetúe la desventaja. Todo demuestra que el problema real no son los genes sino las actitudes que han fomentado las desigualdades en la educación.

¿Qué se puede hacer al respecto? Un camino es combatir de la mejor manera posible la idea de que la matemática es "naturalmente" más complicada para uno que para otro. Si lo es, probablemente se deba a que el alumno ha tenido menos contacto con ella y con los números en general, al haber sido desalentado a asistir a cursos avanzados de matemática o debido a que no ha recibido una buena enseñanza en la materia. Ponerse al día puede ser difícil, pero el hecho de tener que esforzarse más no significa nada con respecto a la potencialidad para aprender estadística.

Para cambiar las ideas equivocadas que pudiéramos tener sobre nuestras propias capacidades, puede ser útil reconocer que existe una creencia errónea y sin valor, ampliamente difundida, que sostiene que la capacidad matemática es innata, algo que se tiene o no (por lo tanto, la conclusión más frecuente es que no tiene sentido estudiar una materia si no existe la menor esperanza de llegar a dominarla). No existe *prueba alguna de la existencia de una capacidad innata* y, sin embargo, sí existe gran cantidad de evidencia que indica que los distintos rendimientos se deben al esfuerzo realizado.

Tobías (1987) cita un estudio comparativo entre estudiantes asiáticos y norteamericanos en un examen internacional de matemática. Los estudiantes norteamericanos fueron superados por completo, pero aún más importante fue el motivo de ese resultado: las entrevistas revelaron que los estudiantes asiáticos consideraban que la capacidad matemática estaba distribuida en forma bastante pareja entre las personas, y pensaban que las diferencias de desempeño eran el resultado del mayor esfuerzo. Los alumnos norteamericanos insistían con que la capacidad para la matemática es un talento raro e innato.

La matemática casi nunca resulta fácil para nadie. Y casi toda persona puede aprender incluso el más complejo de los conceptos matemáticos, si es perseverante y si los conceptos le son bien enseñados. Si para algunas personas la matemática es más sencilla que para otros, sólo se debe a que algunos tienen mayor práctica y experiencia con los números. Pensar que uno ha nacido con menor capacidad para aprender matemática y estadística crea una preocupación adicional que es necesario descartar ahora mismo. Como mencionamos antes, sencillamente no existe prueba alguna que indique diferencias inherentes, y las diferencias de desempeño que en efecto existen no necesariamente predicen algo acerca de uno. Cada uno es un individuo, con su propia capacidad y determinación. Si un alumno necesita trabajar con más esfuerzo para aprobar esta materia, seguramente se sentirá más satisfecho cuando lo logre. Y vale la pena recordar que uno no lo está haciendo sólo por uno mismo, sino que cada mujer y cada persona de color que curse estadística, o cualquier otro curso de matemática a escala universitaria, es en realidad un modelo para aquellos que vendrán después.

Consideremos las palabras pronunciadas por el ex presidente de la Asociación Americana de Matemática:

> La paradoja de nuestros tiempos es que a la vez que la matemática es cada vez más poderosa, sólo los poderosos parecen beneficiarse con ella. La capacidad de pensar matemáticamente, en un sentido amplio, es absolutamente crucial para el desarrollo en prácticamente todas las carreras. La confianza en el manejo de información, el escepticismo en el análisis de argumentos, la perseverancia para penetrar problemas complejos y la capacidad de comunicación sobre temas técnicos en forma oral o escrita, son las artes facultativas que ofrecen las nuevas ciencias matemáticas. (Steen, 1987, p. xviii).

No debemos dejar de aprender estas "artes facultativas" sólo porque alguien nos hizo creer que no podríamos o no querríamos aprenderlas.

TABLAS DE FRECUENCIAS, HISTOGRAMAS Y POLÍGONOS DE FRECUENCIAS SEGÚN SE DESCRIBEN EN PUBLICACIONES CIENTÍFICAS

Los psicológicos investigadores utilizan las tablas de frecuencias, histogramas y polígonos de frecuencias principalmente como pasos previos para análisis estadísticos más elaborados. El estudio realizado por Sanbonmatsu, Posavac y Stasney (1997) aporta un ejemplo de una tabla de frecuencias publicada en un artículo de investigación científica. La investigación analizaba la tendencia de las personas a sobrestimar la probabilidad de que un hecho posible suceda cuando su atención esté puesta en ese hecho. Se informó a los participantes que se estaban examinando cuatro candidatos para un puesto dentro del cuerpo docente de la Universidad de Indiana. Su tarea era estimar la probabilidad de que un candidato en particular fuera contratado. En primer lugar, los participantes hicieron girar una flecha en una ruleta para seleccionar al candidato que evaluarían. Luego se les entregó información sobre el desempeño de los cuatro candidatos. La información que se brindó sobre los candidatos era positiva y, sobre todo, se la preparó procurando que fuera pareja para los cuatro candidatos. Por lo tanto, ninguno de ellos debería haber sido calificado con mayores probabilidades de obtener el empleo con respecto a los otros. Sin embargo, cualquiera fuera el candidato evaluado, los participantes mostraron una firme tendencia a calificar a su propio candidato con mayores probabilidades que los demás. Una de las formas empleadas por Sanbonmatsu (et al., 1997) para mostrar los resultados, consiste en indicar la cantidad de participantes que seleccionaron cada uno, de los nueve enunciados que se les propusieron relativos a la probabilidad de que los candidatos fueran contratados. La tabla 1-6 reproduce la tabla de frecuencias de los participantes. Como puede observarse, muchos más participantes calificaron como "mejor que casual" la probabilidad de que los candidatos que ellos mismos habían evaluado fueran contratados (mayor que 5 en la serie de enunciados).

De todos modos, por lo general, cuando se publican tablas de frecuencias en los artículos, los valores de la variable son categorías y no números. Además, es muy común que en las tablas se utilicen porcentajes de casos más que de cantidades. Por ejemplo, Norcross, Hanych y Terranova (1996) analizaron el Postgrado de Psicología (APA –American Psychologycal Association, Asociación Americana de Psicología–, 1994) e identificaron los cursos exigidos para ser aceptado en la escuela para graduados. La tabla 1-7 fue extraída de esa publicación. La tabla muestra el porcentaje de programas para graduados, la cual exige o prefiere que los aspirantes hayan tomado distintos cursos de psicología. ¡A propósito, podemos observar que un mayor porcentaje (85,25) exige o prefiere que los aspirantes hayan cursado estadística!

En las publicaciones científicas casi nunca se publican histogramas o polígonos de frecuencias (excepto en artículos sobre estadística). En las raras ocasiones en que aparecen, por lo general se presentan en formatos bastante poco comunes. Por ejemplo, Wechsler y sus colegas (1994) realizaron un estudio entre 17.592 estudiantes de 140 facultades de Estados Unidos con respecto a hábitos en la bebida. Una de las preguntas más importantes se refería al porcentaje de estudiantes de las distintas facultades considerados bebedores por diversión. Los investigadores definían

Tabla 1-6.
Frecuencias con que se confirmaron los enunciados sobre la probabilidad de que el candidato asignado fuera contratado: experimento 1.

	Frecuencia ($n = 31$)
1. Es **absolutamente seguro** que (el objetivo) **no** sea contratado	0
2. Las probabilidades de (el objetivo) ser seleccionado para el puesto en el cuerpo docente de la universidad eran **mucho menores** que las de cualquiera de los otros tres candidatos	1
3. Las probabilidades de (el objetivo) ser seleccionado para el puesto en el cuerpo docente de la universidad eran **menores** que las de cualquiera de los otros tres candidatos	3
4. Las probabilidades de (el objetivo) ser seleccionado para el puesto en el cuerpo docente de la universidad eran **apenas menores** que las de cualquiera de los otros tres candidatos	3
5. Las probabilidades de (el objetivo) ser seleccionado para el puesto en el cuerpo docente de la universidad eran **aproximadamente las mismas** que las de cualquiera de los otros tres candidatos	6
6. Las probabilidades de (el objetivo) ser seleccionado para el puesto en el cuerpo docente de la universidad eran **apenas mejores** que las de cualquiera de los otros tres candidatos	8
7. Las probabilidades de (el objetivo) ser seleccionado para el puesto en el cuerpo docente de la universidad eran **mejores** que las de cualquiera de los otros tres candidatos	6
8. Las probabilidades de (el objetivo) ser seleccionado para el puesto en el cuerpo docente de la universidad eran **mucho mejores** que las de cualquiera de los otros tres candidatos	2
9. Es **absolutamente seguro** que (el objetivo) será contratado	2

Fuente: Sanbonmatsu, D. M., Posavac, S.S., & Stasney, R. (1997), tab. 2. "Opiniones subjetivas implícitas en la sobrestimación de probabilidades". *Revista Científica de Psicología Social Experimental [Journal of Experimental Social Psychology]* 33, 276-296. Copyright, 1997, por Academic Press. Reimpreso con autorización.

como bebedores por diversión a aquéllos que informaban haber bebido al menos una vez durante las dos semanas previas a la encuesta (cuatro copas de bebida alcohólica seguidas en el caso de las mujeres y cinco en el caso de los hombres). La figura 1-17 reproduce la tabla que elaboraron. Únicamente una facultad presentó sólo un 1-5% de bebedores por diversión. Sin embargo, bastantes facultades presentaron un 30-50% de estos bebedores. ¡Seis facultades presentaron un 66 -70% de alumnos considerados bebedores por diversión!

Tabla 1-7.
Cursos requeridos como condición previa por programas para graduados que exigen cursos específicos de psicología (*n* = 1.554).

Curso	Exigido	Preferido	Exigido o preferido
Anormal / psicopatología	15,6%	16,9%	32,5%
Niñez / desarrollo	11,2%	24,2%	35,9%
Cognoscitivo	3,6%	9,0%	12,6%
Historia y sistemas	4,9%	12,3%	17,2%
Cursos de laboratorio	7,8%	3,8%	11,6%
Aprendizaje	8,7%	19,6%	28,3%
Fisiológico / biopsicología	5,5%	18,9%	24,4%
Personalidad	12,0%	15,7%	27,7%
Métodos de investigación / diseño experimental	40,0%	26,0%	66,0%
Sensación y percepción	3,1%	7,8%	10,9%
Estadística	56,5%	28,7%	85,2%
Prueba / medición	9,5%	8,0%	17,5%
Social	4,7%	18,7%	23,4%

Fuente: Norcross, J. C., Hanych, J. M., & Terranova, R. D. (1996), tab. 4, Postgrado de Psicología, 1992-1993. *American Psychologist*, 51, 631-643, Copyright, 1996, por la Asociación Americana de Psicología [American Psychological Association]. Reimpreso con autorización.

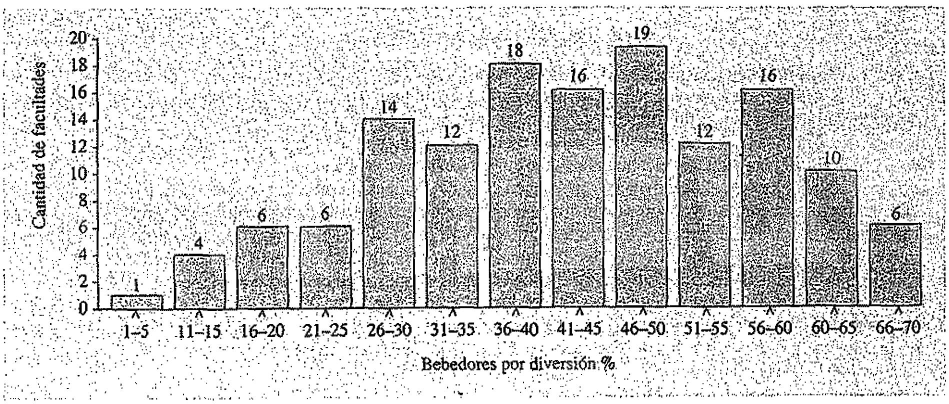


Figura 1-17. Distribución de facultades según porcentaje de bebedores por diversión. (**Fuente**: Wechsler, H., Davenport, A., Dowdall, G., Moeykens, B., & Castillo, S. (7 de diciembre de 1994), tab. 1. "Consecuencias para la salud y el comportamiento provocadas por el hábito de beber por diversión durante la época universitaria: un estudio nacional de alumnos de 140 campus universitarios", *JAMA*, 272, 1674).

Los histogramas y los polígonos de frecuencias sólo en raras ocasiones se publican en artículos de investigación. Es más probable que se vean breves comentarios sobre la forma de la distribución de los registros recolectados durante el estudio, especialmente si la distribución se desvía de lo normal. Speed y Gangestad (1997) aportan una típica descripción, como la mencionada en sus comentarios, sobre un grupo de variables que analizaron: "Estas variables no estaban distribuidas en forma normal (eran positivamente asimétricas)". (p. 930)

Resumen

Los psicólogos utilizan procedimientos de estadística descriptiva para describir, es decir, para resumir y hacer fácilmente comprensibles un grupo de números obtenidos a partir de una investigación.

Un valor es un número o categoría; una variable es una característica que puede tener diferentes valores; una observación es el valor particular correspondiente a una persona en una variable. Con una variable numérica, los valores nos transmiten el grado o cantidad de lo que se mide. Hay dos clases principales de variables numéricas: en el caso de las variables intervalares, los valores representan cantidades iguales de lo que se mide; en el caso de las variables ordinales, los valores sólo representan posiciones relativas. En el caso de las variables nominales, los valores son categorías o nombres.

Una tabla de frecuencias organiza los números en una tabla en la que cada uno de los valores posibles aparecen en una lista a lo largo del margen izquierdo, ordenado de mayor a menor, junto con la cantidad de observaciones que corresponden a cada valor.

Cuando hay una gran cantidad de valores diferentes es más útil construir una tabla de frecuencias agrupadas, que es igual a una tabla de frecuencias común, sólo que las frecuencias se atribuyen a intervalos que incluyen una serie de valores. El tamaño de los intervalos debe elegirse de tal modo que (a) la cantidad total de intervalos sea de entre 5 y 15; (b) sea un número común, simple, y (c) el límite inferior de cada intervalo sea múltiplo del tamaño del intervalo.

El patrón de las frecuencias puede representarse con un histograma, es decir, una especie de gráfico en el que la altura de cada barra es la frecuencia para un valor o intervalo determinado, y en el que no existen espacios entre las barras. Los polígonos de frecuencias son otra alternativa de los histogramas; en ellos, una línea conecta puntos, es decir, la altura de cada uno de los cuales representa la frecuencia de un valor o intervalo determinado.

La forma general del histograma o polígono de frecuencias puede ser unimodal (con un solo pico), bimodal, multimodal (que incluye al bimodal), o rectangular (sin picos); puede ser simétrica o asimétrica (con una larga cola) hacia la derecha o hacia la izquierda; y con respecto a la curva normal con forma de campana, puede presentar curtosis (con colas que son muy anchas o muy angostas).

A veces se puede distorsionar la representación gráfica de información para el público en general, de tal manera que a simple vista resulte engañosa, como por ejemplo, utilizando intervalos que no son iguales o exagerando las proporciones.

Las tablas de frecuencias, los histogramas y los polígonos de frecuencias rara vez aparecen en publicaciones científicas. Cuando aparecen, por lo general lo hacen en formatos no tradicionales o presentando frecuencias (o porcentajes) para varias categorías, más que para los diferentes valores numéricos de una variable.

Términos clave

- Bimodal.
- Efecto techo.
- Estadística descriptiva.
- Intervalar.
- Efecto piso.
- Distribución de frecuencias.
- Polígono de frecuencias.
- Tabla de frecuencias.
- Tabla de frecuencias agrupadas.
- Histogramas.
- Estadística inferencial.
- Intervalo.
- Curtosis.
- Niveles de medición.
- Multimodal.
- Variable nominal.
- Curva normal.
- Variable numérica.
- Variable ordinal.
- Rectangular.
- Valor observado u observación.
- Asimétrico.
- Simétrico.
- Unimodal.
- Valor.
- Variable.

Ejercicios

Los ejercicios implican la realización de cálculos o tabulaciones. La mayoría de los problemas estadísticos reales se resuelven por computadora, pero aun así, es conveniente realizar estos ejercicios manualmente para incorporar el método de trabajo.

Para adquirir práctica en la utilización de una computadora, para resolver problemas estadísticos, se puede utilizar la sección de computación de cada capítulo, publicada en la *Guía de estudio y libro de tareas de computación para el alumno [Student's Study Guide and Computer Workbook]* que acompaña este libro.

Todos los datos de esta sección son ficticios (a menos que se especifique lo contrario). Las respuestas a los ejercicios de la serie I se encuentran al final del libro.

SERIE I

1. Indique el nivel de medición de cada una de las siguientes variables: a) grupo étnico al que pertenece una persona, b) cantidad de veces que un animal equivoca el camino en un laberinto y c) posición en la que uno finaliza una carrera.

2. A continuación, aparecen las observaciones de una medición de sensibilidad olfativa realizada a un grupo de chefs asistentes a un congreso nacional:

96, 83, 59, 64, 73, 74, 80, 68, 87, 67, 64, 92, 76, 71, 68, 50, 85, 75, 81, 70, 76, 91, 69, 83, 75

Confeccione a) una tabla de frecuencias, b) una tabla de frecuencias agrupadas y c) un histograma de las frecuencias agrupadas; y d) describa la forma general de la distribución.

3. Las observaciones que aparecen a continuación representan la cantidad de minutos que tardó cada uno de los integrantes de un grupo de niños de 10 años de edad en completar una serie de rompecabezas abstractos:

24, 83, 36, 22, 81, 39, 60, 62, 38, 66, 38, 36, 45, 20, 20, 67, 41, 87, 41, 82, 35, 82, 28, 80, 80, 68, 40, 27, 43, 80, 31, 89, 83, 24

Confeccione a) una tabla de frecuencias agrupadas y b) un histograma que represente la tabla.

4. Supongamos que se consultó a 50 alumnos acerca de la cantidad de horas que estudiaron durante este fin de semana, y que los mismos dieron las siguientes respuestas:

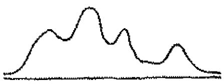
11, 2, 0, 13, 5, 7, 1, 8, 12, 11, 7, 8, 9, 10, 7, 4, 6, 10, 4, 7, 8, 6, 7, 10, 7, 3, 11, 18, 2, 9, 7, 3, 8, 7, 3, 13, 9, 8, 7, 7, 10, 4, 15, 3, 5, 6, 9, 7, 10, 6

Confeccione a) una tabla de frecuencias, b) una tabla de frecuencias agrupadas y c) un polígono de frecuencias que represente la tabla de frecuencias agrupadas, y d) describa la forma general de la distribución.

5. Describa las formas de las tres distribuciones ilustradas.

(a)

(b)



(c)

6. Dibuje un ejemplo de cada una de las siguientes distribuciones: a) simétrica, b) rectangular y c) asimétrica hacia la derecha.

7. Explique a alguien que nunca ha asistido a un curso sobre estadística el significado de los siguientes términos: a) distribución simétrica, unimodal y b) distribución unimodal negativamente asimétrica. (Asegúrese de explicar también en su primera respuesta el significado de distribución).

SERIE II

1. Explique y ejemplifique cada uno de los siguientes tipos de variable: a) intervalar, b) ordinal y c) nominal.

2. A continuación, aparecen las velocidades de automóviles cronometrados por radar durante una tarde en una ruta de una zona con límite de velocidad de 35 millas por hora:

30, 36, 42, 36, 30, 52, 36, 34, 36, 33, 30, 32, 35, 32, 37, 34, 36, 31, 35, 20, 24, 46, 23, 31, 32, 45, 34, 37, 28, 40, 34, 38, 40, 52, 31, 33, 15, 27, 36, 40

Confeccione a) una tabla de frecuencias, b) una tabla de frecuencias agrupadas, c) un histograma de las frecuencias agrupadas y d) un polígono de frecuencias de las frecuencias agrupadas, y e) describa la forma general de la distribución.

3. Las siguientes son las cantidades de regalos adquiridos por 25 familias entrevistadas al azar en un centro de compras local a fines de la temporada de vacaciones:

22, 18, 22, 26, 19, 14, 23, 27, 2, 18, 28, 28, 11, 16, 34, 28, 13, 21, 32, 17, 6, 29, 23, 22, 19

Confeccione a) una tabla de frecuencias agrupadas y b) un polígono de frecuencias con las frecuencias agrupadas, y c) describa la forma general de la distribución.

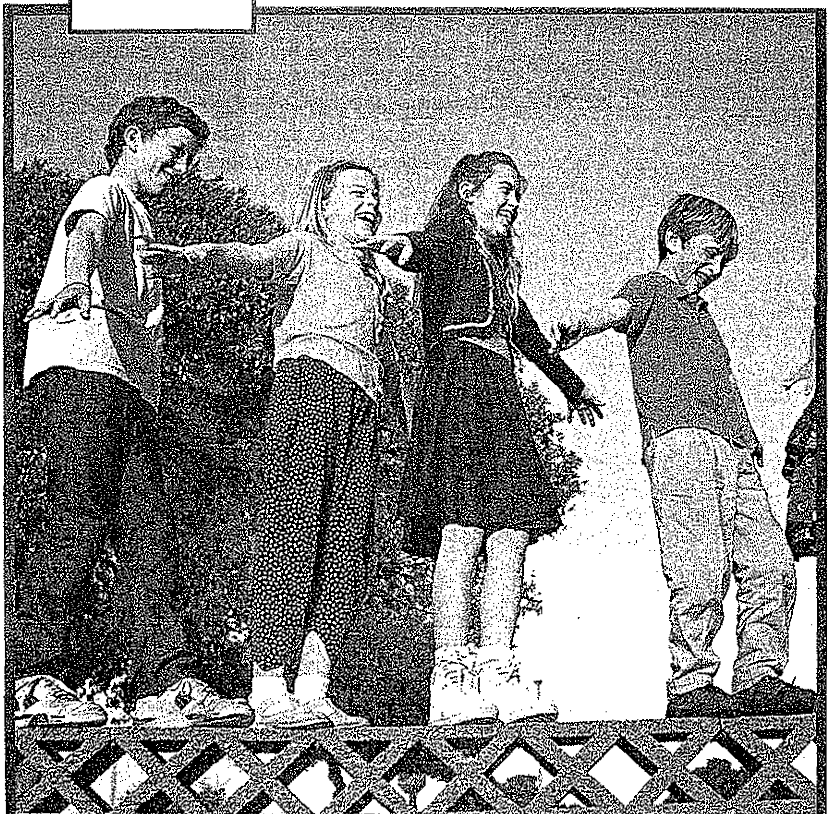
4. Elija un libro y una página del mismo (escoja una página con al menos 30 renglones). Confeccione una lista de la cantidad de palabras en cada renglón; luego utilice esa lista como información. Confeccione a) una tabla de frecuencias, b) una tabla de frecuencias agrupadas, c) un histograma de las frecuencias agrupadas y d) un polígono de frecuencias con las frecuencias agrupadas, y e) describa la forma general de la distribución.

5. Explique a una persona que nunca ha asistido a un curso sobre estadística el significado de a) tabla de frecuencias agrupadas y b) histograma.

6. Cree un ejemplo que presente las siguientes formas de distribución: a) bimodal, b) prácticamente rectangular y c) positivamente asimétrica. No utilice ejemplos dados en este libro o en clase.

7. Busque en periódicos o revistas un gráfico cuyo aspecto resulte engañoso a causa de la utilización de intervalos desiguales o de la exageración de las proporciones.

2 Media, varianza, desvío estándar y puntuaciones Z



Descripción del capítulo

- Media.
- Medidas alternativas de la tendencia central.
- Varianza y desvío estándar.
- Puntuaciones Z.
- Controversias y limitaciones: la tiranía de la media.
- La media y el desvío estándar según se describen en publicaciones científicas.
- Resumen.
- Términos clave.
- Ejercicios.
- Apéndice del capítulo: fórmulas optativas para el cálculo de la varianza y el desvío estándar.

Como señalamos en el capítulo 1, el propósito de la estadística descriptiva es hacer fácilmente comprensibles un grupo de observaciones. Hemos visto algunas formas de lograr esa comprensión a través de tablas y gráficos. Es este capítulo, consideraremos las principales técnicas estadísticas para describir un grupo de observaciones utilizando ciertos números. Estos números son: la media, la varianza, el desvío estándar y las puntuaciones Z. La media es el promedio. La varianza y el desvío estándar describen el grado de variación de las observaciones. Una puntuación Z describe la desviación de una observación en particular respecto del promedio.

MEDIA

Comúnmente, el mejor número para describir un grupo de observaciones es el promedio normal, es decir, la suma de todas las observaciones dividida por la cantidad de observaciones. En estadística, ese promedio se denomina **media**. A veces se dice que el promedio o media de un grupo de registros muestra la **tendencia central** o el valor típico o representativo de un grupo de observaciones. Más adelante veremos que existen otras formas, además de la media, para describir la tendencia central de un grupo de observaciones.

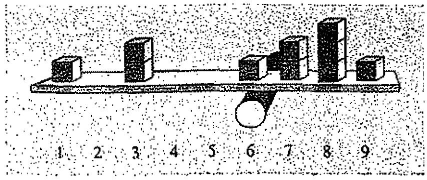
Supongamos que una psicoterapeuta observó cuántas sesiones habían tardado sus últimos 10 pacientes en completar una terapia breve. Las cantidades de sesiones eran las siguientes:

7, 8, 8, 7, 3, 1, 6, 9, 3, 8

La media de las 10 observaciones anteriores es 6 (la suma de 60 sesiones dividida por 10 pacientes). Es decir, en promedio, los últimos 10 pacientes de la terapeuta habían asistido a 6 sesiones. Así, la información referida a los 10 pacientes se resume sólo en este número.

A muchos estudiantes les resulta útil visualizar la media como una especie de punto de equilibrio de la distribución de observaciones. Intentemos visualizar una tabla en equilibrio sobre un tronco, como un

Figura 2-1. Media de la distribución de cantidad de sesiones de terapia realizadas según un ejemplo ficticio, ilustrada a través de una analogía con cubos apoyados encima de una tabla en equilibrio sobre un tronco.



sube y baja rudimentario. Imaginemos pilas de cubos distribuidos a lo largo de la tabla según los valores que representan, es decir, un cubo para cada observación de la distribución. (La figura resultante es similar a un histograma construido con cubos). La media sería el punto de la tabla donde el peso de los cubos se equilibra perfectamente. La figura 2-1 representa lo antedicho utilizando el ejemplo de la cantidad de sesiones a las que asistieron los 10 pacientes de nuestra terapeuta imaginaria.

La figura 2-2 muestra algunos otros ejemplos. Cabe destacar, que ni siquiera es necesario que haya un cubo exactamente en el punto de equilibrio. Es decir, la media no necesariamente debe corresponder a una observación real en la distribución. La media es simplemente el promedio de las observaciones, el punto de equilibrio. La media incluso podría ser un número cuya aparición en la distribución fuera imposible, como en el caso de una media representada por un número decimal cuando todos los números en la distribución deben ser números enteros (2,3 niños, por ejemplo). Otra característica es que los cubos pueden estar muy separados o muy juntos y que no necesiten estar distribuidos en forma pareja. En cualquiera de esos casos, aun es posible encontrar un punto de equilibrio. (Cabe mencionar que esta analogía, que utiliza cubos en equilibrio encima de una tabla apoyada sobre un tronco, funcionaría en la realidad sólo si la tabla no tuviera peso).

Fórmula para obtención de la media y símbolos estadísticos

La regla para el cálculo de la media consiste en sumar todas las observaciones y dividirlas por la cantidad de las mismas. Se expresa a través de la siguiente fórmula:

$$M = \frac{\Sigma X}{N} \tag{2-1}$$

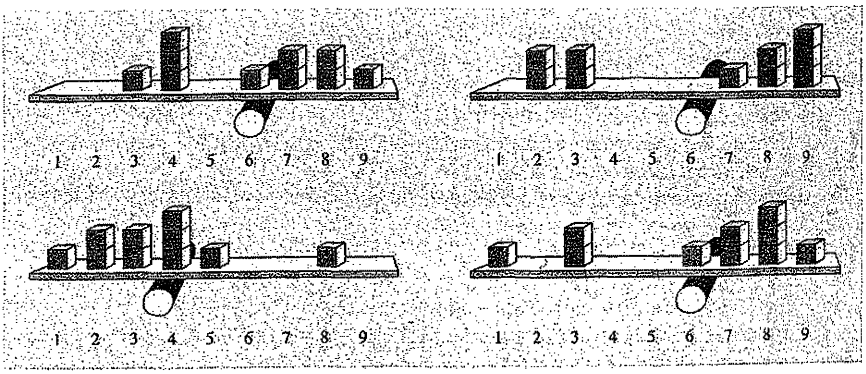


Figura 2-2. Medias de varias distribuciones ficticias ilustradas utilizando la analogía de los cubos apoyados encima de una tabla en equilibrio sobre un tronco.

M es un símbolo que representa la media. (Más adelante aprenderemos otro símbolo para representar la media, la letra griega μ "mu", que se utiliza en circunstancias particulares. También es bastante utilizado un tercer símbolo, $\bar{X}$, a veces denominado X-raya).

Σ, la letra griega mayúscula "sigma," es el símbolo que representa la "suma de"; significa "suma de todas las cantidades siguientes". Es el símbolo aritmético especial más comúnmente utilizado en estadística.

La X se refiere a las observaciones en la distribución de la variable X. Podríamos haber elegido cualquier otra letra. Sin embargo, cuando existe sólo una distribución, generalmente se la denomina X. En capítulos posteriores veremos situaciones en las que se analizan dos distribuciones al mismo tiempo. En ese caso, se utiliza una segunda letra, generalmente la Y. Otra alternativa es utilizar subíndices, como por ejemplo X_1 y X_2. En el caso de un tratamiento matemático más formal de la estadística, los símbolos utilizados en varias fórmulas son aún más complejos. Es precisamente esa complejidad la que permite que las fórmulas representen casos complicados sin confusión. Sin embargo, los libros de estadística para psicólogos, aun los textos más avanzados, utilizan símbolos simples. La forma más simple rara vez crea ambigüedad en las fórmulas estadísticas que utilizan los psicólogos.

ΣX significa "la suma de X". Indica que se deben sumar todos los valores observados de la distribución de la variable X. Supongamos que X se refiere a la cantidad de sesiones de terapia en la distribución de nuestro ejemplo. ΣX sería igual a 60, la suma de $7 + 8 + 8 + 7 + 3 + 1 + 6 + 9 + 3 + 8$.

N es un número. Se utiliza en estadística para indicar la cantidad de observaciones de una distribución. En nuestro ejemplo existen 10 observaciones, por lo tanto, N es igual a 10.

Resumiendo, la fórmula indica dividir la suma de todas las observaciones de la distribución de la variable X por la cantidad total de observaciones N. En nuestro ejemplo, significa que debemos dividir 60 por 10. La fórmula sería la siguiente:

$$M = \frac{\Sigma X}{N} = \frac{60}{10} = 6$$

Otro ejemplo de cálculo de la media

Analicemos los ejemplos del capítulo 1. Las puntuaciones de estrés (Aron et al., 1995) fueron las siguientes:

4, 7, 7, 7, 8, 8, 7, 8, 9, 4, 7, 3, 6, 9, 10, 5, 7, 10, 6, 8, 7, 8, 7, 8, 7, 4, 5, 10, 10, 0, 9, 8, 3, 7, 9, 7, 9, 5, 8, 5, 0, 4, 6, 6, 7, 5, 3, 2, 8, 5, 10, 9, 10, 6, 4, 8, 8, 8, 4, 8, 7, 3, 8, 8, 8, 8, 7, 9, 7, 5, 6, 3 ,4, 8, 7, 5, 7, 3, 3, 6, 5, 7, 5, 7, 8, 8, 7, 10, 5, 4, 3, 7, 6, 3, 9, 7, 8, 5, 7, 9, 9, 3, 1, 8, 6, 6, 4, 8, 5, 10, 4, 8, 10, 5, 5, 4, 9, 4, 7, 7, 7, 6, 6, 4, 4, 4, 9, 7, 10, 4, 7, 5, 10, 7, 9, 2, 7, 5, 9, 10, 3, 7, 2, 5, 9, 8, 10, 10, 6, 8, 3

Calculando la media se puede resumir toda esta información en un sólo número. La media se calcula sumando todas las puntuaciones de estrés y dividiendo el resultado por la cantidad de puntuaciones de estrés. Es decir, se suman las puntuaciones de estrés, $4 + 7 + 7 + 7 + 8 + 8$, y así sucesivamente, obteniendo un total de 975. Luego se divide el total por 151, es decir, la cantidad de observaciones.

$$M = \frac{\Sigma X}{N} = \frac{975}{151} = 6{,}46$$

La fórmula nos indica que la puntuación promedio de estrés en la escala de 10 puntos fue de 6,46 (redondeando). Esta cifra se encuentra claramente por encima del punto medio de la escala. El ejemplo también puede representarse gráficamente. Consideremos otra vez el histograma como una pila de cubos encima de una tabla, y la media 6,46 como el punto en el que la tabla se equilibra sobre el fulcro que tiene debajo. (*véase* figura 2-3). Este único número simplifica enormemente la información de las 151 puntuaciones de estrés.

De modo similar, analicemos el ejemplo de las interacciones sociales de los alumnos (McLaughlin-Volpe et al., 1998). Las cantidades de interacciones de los 94 alumnos durante una semana fueron las siguientes:

48, 15, 33, 3, 21, 19, 17, 16, 44, 25, 30, 3, 5, 9, 35, 32, 26, 13, 14, 14, 47, 47, 29, 18, 11, 5, 19, 24, 17, 6, 25, 8, 18, 29, 1, 18, 22, 3, 22, 29, 2, 6, 10, 29, 10, 21, 38, 41, 16, 17, 8, 40, 8, 10, 18, 7, 4, 4, 8, 11, 3, 23, 10, 19, 21, 13, 12, 10, 4, 17, 11, 21, 9, 8, 7, 5, 3, 22, 14, 25, 4, 11, 10, 18, 1, 28, 27, 19, 24, 35, 9, 30, 8, 26

La tabla de frecuencias, la tabla de frecuencias agrupadas, el histograma y el polígono de frecuencias que construimos en el capítulo 1 simplificaron considerablemente la visualización de los datos. Pero incluso, después de todo ese proceso también sería útil obtener un resumen de un sólo número. Por lo tanto, podemos calcular la media en la forma usual. En este caso:

$$M = \frac{\Sigma X}{N} = \frac{1.635}{94} = 17{,}40$$

Es decir, si sumamos las cantidades de interacciones de los 94 alumnos, la "suma de X" da 1.635. Al dividir este número por la cantidad de observaciones, obtenemos una media de interacciones de 17,40. La figura 2-4 grafica este caso.

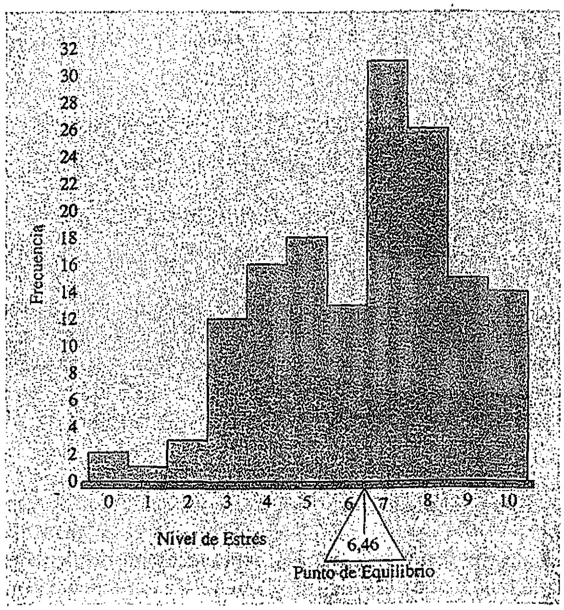


Figura 2-3. Analogía de cubos apoyados encima de una tabla en equilibrio sobre un punto de apoyo (utilizando un histograma) que representa la media de las puntuaciones del nivel de estrés dadas por 151 estudiantes de estadística. (**Fuente**: Aron, Paris, & Aron, 1995).

Otro de los principales ejemplos analizados en el capítulo 1 se refería al tiempo utilizado para leer oraciones ambiguas. La media se calcula en la forma usual:

$$M = \frac{\Sigma X}{N} = \frac{275{,}5}{100} = 2{,}755$$

La figura 2-5 representa este caso gráficamente.

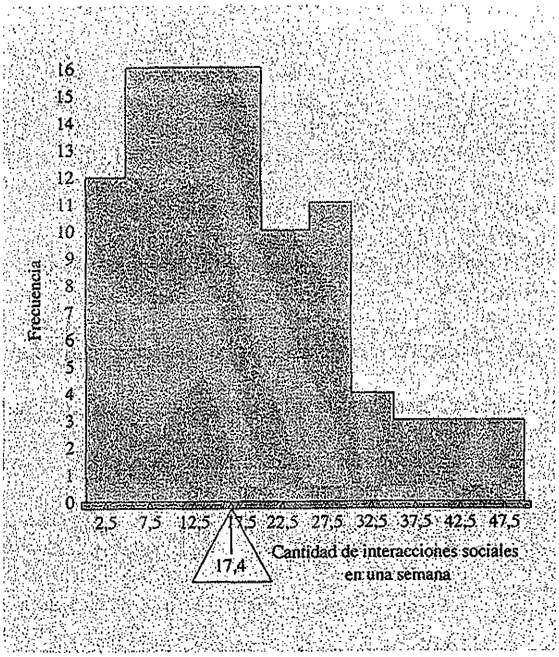


Figura 2-4. Analogía de cubos apoyados encima de una tabla en equilibrio sobre un punto de apoyo (utilizando un histograma) que representa la media de la cantidad de interacciones sociales vividas por 94 alumnos universitarios durante una semana (**Fuente:** McLaughlin-Volpe et al., 1998).

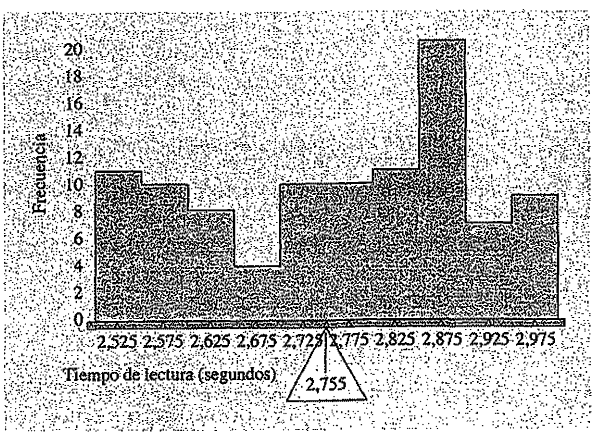


Figura 2-5. Analogía de cubos apoyados encima de una tabla en equilibrio sobre un punto de apoyo (utilizando un histograma) que representa la media de una distribución ficticia del tiempo utilizado por 100 sujetos para leer oraciones ambiguas (en segundos).

MEDIDAS ALTERNATIVAS DE TENDENCIA CENTRAL

La media es sólo uno de los varios medios descriptivos de la tendencia central, el valor típico o representativo. Otra alternativa es la **moda**. La moda es el valor observado más común en una distribución. En nuestro ejemplo sobre las sesiones de terapia, la moda es 8, porque existen tres pacientes que asistieron a 8 sesiones y no existe ninguna otra cantidad de sesiones que haya sido completada por igual o mayor cantidad de pacientes. La moda también puede considerarse el valor con mayor frecuencia en una tabla de frecuencias, o bien el punto alto o pico de la distribución en un polígono de frecuencias o en un histograma (tal como lo ilustra la figura 2-6).

En una distribución perfectamente simétrica y unimodal, la moda coincide con la media. ¿Qué ocurre cuando la media y la moda no coinciden? En ese caso, la moda coincide en menor grado que la media, por lo cual, en líneas generales, consideraríamos un valor bien representativo de la distribución. Además, es posible cambiar algunas de las observaciones de una distribución (*véase* la figura 2-7) sin afectar la moda, mientras que prácticamente cualquier cambio que se realice afecta la media. *Por lo tanto, la media es más representativa* de todas las observaciones de una distribución. Por estas y otras razones, los psicólogos rara vez utilizan la moda.

Otra alternativa de la media es la **mediana**. Si ordenamos todas las observaciones de menor a mayor, el valor del medio es la mediana. La figura 2-8 muestra los registros correspondientes a la cantidad de sesiones de terapia, ordenados de menor a mayor. En este ejemplo, el quinto y sexto caso (los dos del medio) son iguales a 7. Es decir, en cualquiera de los dos casos, la mediana es 7. Uno de los errores más comunes que cometen los estudiantes de estadística al calcular la mediana es *no ordenar primero* las observaciones de menor a mayor.

Cuando existe una cantidad par de casos, la mediana puede estar entre dos números diferentes. En el ejemplo anterior, teníamos una cantidad par de registros, pero los dos casos del medio presentaban el mismo número; por lo tanto, no surgía ningún inconveniente. Cuando los dos casos medios son diferentes, se utiliza el promedio de los dos. En el ejemplo del tiempo requerido para leer oraciones ambiguas hay exactamente 100 casos. El caso 50 (de menor a mayor) es de 2,76, y el 51 de 2,78. La mediana es, por lo tanto, 2,77; el promedio de 2,76 y 2,78.

En ciertos casos, la mediana indica con más precisión que la media la tendencia central de un grupo de observaciones. Esto sucede cuando existen unas pocas observaciones extremas que afectarían notoriamente la media pero no influirían en la mediana. Por ejemplo, supongamos que de 100 familias que trabajan en una plantación de bananas en América Central, 99 obtienen un ingreso anual de \$100 y 1 (la familia del propietario) obtiene un ingreso anual de \$90.100. La media del ingreso familiar en esta plantación sería de \$1000 (99 x 100 = 9.900; 9.900 + 90.100 = 100.000; 100.000/100 = 1.000). Sin embargo, ninguna de las familias obtiene un ingreso siquiera cercano a los \$1.000, por lo que esta cantidad resulta ser completamente engañosa. En este caso, la mediana del ingreso por familia sería de \$100, una cifra mucho más representativa de cualquier persona a la que uno pudiera acercarse por azar en la plantación.

Los resultados del tiempo de reacción son otro ejemplo de los casos en los que podría ser preferible la mediana. Supongamos que los tiempos en cinco pruebas fueron (en segundos) 0,74, 0,86, 2,32, 0,79 y 0,81. La marca de 2,32 segundos puede haber ocurrido porque la persona se distrajo momentáneamente. Por lo tanto, podría ser mejor utilizar la mediana para describir la tendencia central, ya que esto le quitaría influencia a la única observación extrema, y probablemente sea lo correcto.

La importancia del indicador de tendencia central utilizado se refleja en una reciente controversia entre psicólogos que estudian la base evolutiva de la elección de la pareja humana. Una serie de teóricos (p. ej., Buss & Schmitt, 1993) sostienen que a lo largo de sus vidas los hombres preferirían tener muchas más parejas que las mujeres. Según esta visión, la evolución sería la cau-

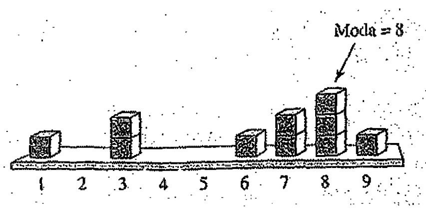


Figura 2-6. Representación gráfica de la moda como *punto más alto* en el histograma de una distribución, utilizando el ejemplo ficticio de la cantidad de sesiones de terapia tomadas por 10 pacientes.

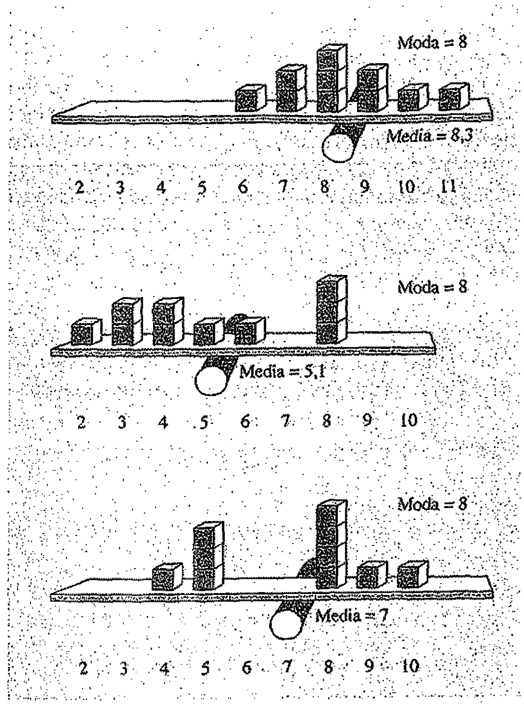


Figura 2-7. Representación gráfica del efecto causado en la media y en la moda por el cambio de algunos valores, utilizando el ejemplo ficticio de la cantidad de sesiones de terapia tomadas por 10 pacientes.

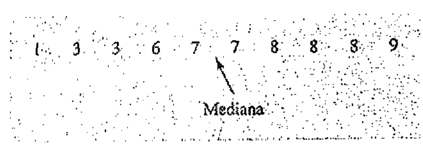


Figura 2-8. Representación gráfica de la mediana como el valor del medio al ordenar los registros de menor a mayor, utilizando el ejemplo ficticio de la cantidad de sesiones de terapia tomadas por 10 pacientes.

sa de que las mujeres prefieran tener sólo una pareja estable, debido a que una mujer sólo puede tener una pequeña cantidad de hijos durante su vida, y es más probable que los genes de la mujer sobrevivan si esos pocos hijos son bien cuidados. Los hombres, sin embargo, pueden tener una gran cantidad de hijos durante su vida; por eso, según la misma teoría, para ellos lo mejor es una postura semejante al disparo de escopeta. Si tienen muchas parejas es más probable que sus genes sobrevivan. Coherentemente con esta presunción, los psicólogos evolucionistas descubrieron que los hombres expresaban necesitar muchas más parejas que las mujeres.

Otros teóricos (p.ej., Miller & Fishkin, 1998), sin embargo, han cuestionado esta visión. Sostienen que hombres y mujeres preferirían aproximadamente la misma cantidad de parejas debido a que los individuos que tienen una predisposición básica a buscar un lazo íntimo fuerte son los que tienen las mayores probabilidades de sobrevivir a la niñez, y que este deseo de lazos fuertes perdura (y tiene otros beneficios) en la etapa adulta. Los mismos investigadores también preguntaron a mujeres y hombres cuántas parejas necesitaban, y los resultados mostraron la misma diferencia en cuanto a las medias; los hombres necesitaban un promedio de 64,32 y las mujeres un promedio de 2,79. Sin embargo, la escena cambia drásticamente si observamos la mediana o la moda (*véase* tabla 2-1). La figura 2-9, tomada directamente de la publicación preparada por los investigadores, nos explica la situación. La mayoría de las mujeres y los hombres desean sólo una pareja; *unos pocos desean más* de una, y *algunos desean muchas más*. La gran diferencia reside en que hay muchos más hombres dentro del pequeño grupo que desea muchas más parejas. (Los valores observados más extremos estaban tan alejados –los hombres que deseaban más de 100 parejas–, que los investigadores ni siquiera los incluyeron al calcular las medias).

Por lo tanto, ¿cuál de las dos teorías es la correcta? Tal vez uno podría sostener cualquiera de las dos formas para analizar esta información. La verdad es que concentrarse sólo en la media, en este caso, desfigura drásticamente la realidad de la distribución.

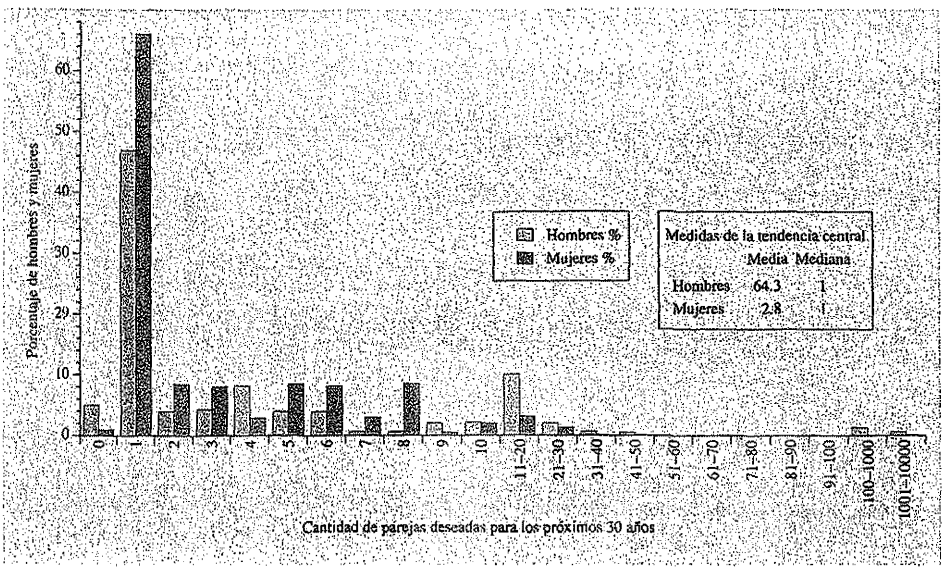


Figura 2-9. Distribuciones de la cantidad ideal de parejas deseadas para un período de 30 años por hombres y mujeres.
Nota: Con el fin de incluir todos los datos, juntamos las categorías que se encontraban más alejadas a lo largo de la cola de estas distribuciones. Si cada categoría representara sólo un número, sería más evidente que la cola es muy chata y que las distribuciones son aún más asimétricas de lo que aquí parece. [**Fuente:** Miller, L. C., & Fishkin, S. A. (1997), fig. 8-1. "Sobre la dinámica del enlace humano y el éxito reproductivo: buscando ventanas en la interfase ambiental adaptada a humanos". En: *J. A. Simpson & D. T. Kendrick (Eds.), Psicología Social Evolutiva [Evolutionary Social Psychology]*. Mahwah, NJ: Lawrence Erlbaum Associates, Inc., Editores.]

De un modo más general, la mediana (y a veces la moda) se utiliza ocasionalmente en psicología como estadística descriptiva. Como hemos visto, es más probable que se utilice en situaciones en las que existen algunos registros extremos que harían que la media no represente el cuerpo principal de casos. También existen circunstancias en las que los psicólogos utilizan la mediana como parte de técnicas estadísticas más complejas.

Tabla 2-1.
Respuestas de 106 hombres y 160 mujeres a la pregunta: "¿Cuántas parejas desearía tener en los próximos 30 años?"

	Media	Mediana	Moda
Mujeres	2,8	1	1
Hombres	64,3	1	1

Fuente: Miller & Fishkin, 1997.

Sin embargo, a menos que existan valores extremos, los psicólogos casi siempre utilizan la media como medida de la tendencia central. En realidad, la media cumple la función de piedra angular para la mayoría de las otras técnicas estadísticas.

VARIANZA Y DESVÍO ESTÁNDAR

Además de la tendencia central, los investigadores necesitan conocer la dispersión de una distribución. Por ejemplo, supongamos que alguien preguntara: "¿Cuántos años tienen los alumnos de determinada clase de estadística?" En una universidad ubicada en una ciudad con muchos estudiantes que retoman los estudios o estudian por horas, la edad media podría llegar a ser 38. Uno podría decir a quien le preguntara: "La edad promedio es 38". Pero esto no reflejaría toda la situación. Sería posible, por ejemplo, tener una media de 38 porque cada estudiante en la clase tiene exactamente 38 años de edad. O podríamos tener una media de 38 porque la mitad de la clase tiene 18 años y la otra mitad 58. Serían dos situaciones muy distintas.

La figura 2-10 representa tres distribuciones de frecuencias diferentes, con la misma media pero con diferentes grados de dispersión de las observaciones alrededor de la media; y otras tres con diferentes medias pero con el mismo grado de dispersión.[1]

Varianza

La **varianza** de un grupo de observaciones indica la dispersión de esos valores alrededor de la media. Para ser más precisos, la varianza es el promedio de los cuadrados de la diferencia entre cada observación y la media. A continuación, detallamos los pasos para calcular la varianza:

[1] Esta sección está dedicada a la varianza y al desvío estándar como indicadores de dispersión. Existe otra forma de describir la dispersión de un grupo de observaciones, la **amplitud**, el registro mayor menos el registro menor. Supongamos que en una clase en particular el registro más alto en un examen parcial es 98 y el menor es 60; la amplitud es 38 (es decir 98 – 60 = 38). La amplitud rara vez es utilizada por investigadores psicológicos ya que se trata de un medio muy burdo de describir la dispersión. Es burdo debido a que no tiene en cuenta la distancia entre las observaciones dentro de la distribución.

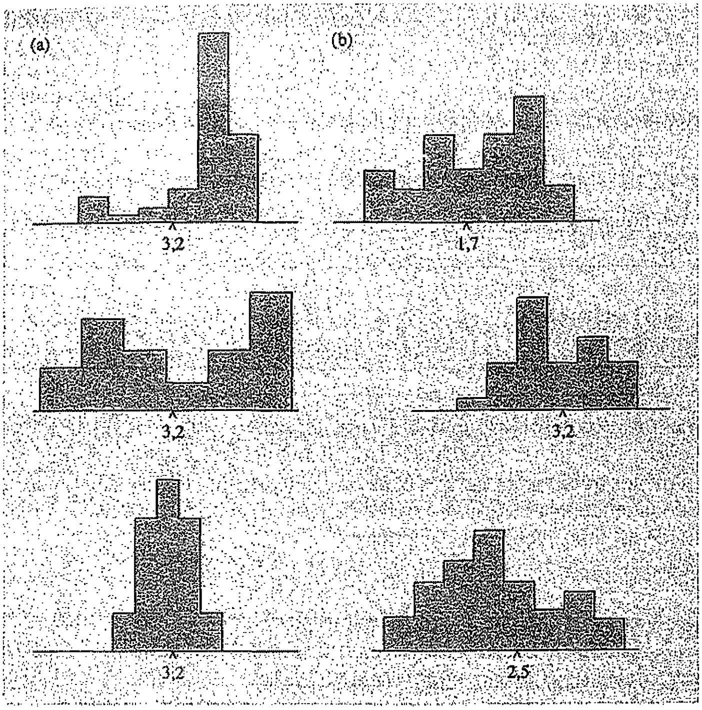


Figura 2-10. Ejemplos de distribuciones con (a) la misma media con diferentes grados de dispersión y (b) diferentes medias con el mismo grado de dispersión.

1. Restar la media a cada observación para obtener el **desvío** de cada una de ellos. El desvío indica la distancia entre la observación en cuestión y la media.

2. Elevar cada uno de los desvíos al cuadrado (multiplicar cada uno de ellos por sí mismo). Se obtiene así el desvío cuadrático de cada registro.

3. Sumar los desvíos cuadráticos. El total logrado con este cálculo se denomina **suma de los cuadrados.**

4. Dividir la suma de los cuadrados por la cantidad de desvíos cuadráticos (es decir, por la cantidad de observaciones). Se obtiene así el promedio o media de desvíos cuadráticos, es decir, la varianza.

Aunque este procedimiento pueda parecer un poco extraño o difícil de recordar al principio, en verdad funciona muy bien. Supongamos que una distribución es más dispersa que otra. La distribución con mayor dispersión presenta una varianza mayor porque la misma dispersión hace que los desvíos sean mayores. Si los desvíos son mayores, los desvíos cuadráticos también lo son y, por lo tanto, también la varianza.

En el ejemplo de la clase en la que todos tenían 38 años de edad, la varianza sería exactamente 0. Es decir, no habría varianza. (En términos numéricos, el desvío de cada persona sería 38 – 38 = 0; 0 al cuadrado es 0. El promedio de 0 es 0). Por el contrario, la clase con la mitad de los alumnos de 18 años de edad y la otra mitad de 58 años de edad tendría una varianza bastante alta, es decir, 400. (Los alumnos de 18 años de edad tendrían cada uno un desvío de 18 – 38 = –20. Los alumnos de 58 años de edad tendrían desvíos de 58 – 38 = 20. En ambos casos, los desvíos cuadráticos, tanto –20 al cuadrado como 20 al cuadrado, darían como resultado 400. Y, cuando todos los números son 400, el promedio es 400).

La varianza es importante en muchos otros procedimientos estadísticos (incluso en la mayoría de los temas tratados en la segunda mitad de este libro). Sin embargo, la varianza se utiliza sólo ocasionalmente como estadística descriptiva, debido a que está basada en desvíos cuadráticos, y los desvíos cuadráticos no transmiten claramente la dispersión de las observaciones. Son verdaderos desvíos o son no cuadráticos. Por ejemplo, queda claro que una clase con una varianza de 400 presenta una distribución mucho más dispersa que otra cuya varianza es 200. Sin embargo, el número 400 no refleja con claridad la variación real entre las edades, ninguna de las cuales se acerca a 400.[2]

Desvío estándar

La estadística más ampliamente utilizada para describir la dispersión de una distribución es el **desvío estándar**. El desvío estándar es la raíz cuadrada positiva de la varianza: para encontrar el desvío estándar, primero es necesario calcular la varianza y luego sacar su raíz cuadrada. Si la varianza de una distribución es 400, el desvío estándar es 20; si la varianza es 9, el desvío estándar es 3, y si la varianza es 100, el desvío estándar es 10.

La varianza se basa en los desvíos de la media, al cuadrado. Por lo tanto, su raíz cuadrada, el desvío estándar, se basa en la distancia simple, no elevada al cuadrado, de la media. **Sin entrar en detalles, el desvío estándar es el promedio de las diferencias entre las observaciones y la media.** Por ejemplo, analicemos una clase donde las edades presentan un desvío estándar de 20 años. Esto nos indicaría que las edades se dispersan, en promedio, aproximadamente 20 años en cada dirección a partir de la media. Conocer el desvío estándar ofrece una idea general del grado de dispersión.

Daremos otro ejemplo. La distribución de la cantidad de hijos por familia en un país en particular podría tener una media de 4 y un desvío estándar de 1. Significaría que, por cada familia con exactamente cuatro hijos (desvío 0 de la media), bien podríamos encontrar una con seis o dos hijos (desvío de 2 hijos de la media). Sin embargo, podría no funcionar de esa forma. Podría ser que la mitad de las familias tuvieran exactamente 5 y la otra mitad exactamente 3. O podría ser que la mayoría tuviera 4, pero unas pocas no tuvieran ninguno y otras pocas tuvieran 8 (*véase* figura 2-11). No obstante, conocer el desvío estándar brinda una noción general del grado de dispersión, aun cuando no indique la forma precisa de distribución.

El desvío estándar no es **exactamente** el promedio de las diferencias entre las observaciones y la media. Para ser precisos, el desvío estándar es la raíz cuadrada del promedio de los desvíos cuadráticos de la media. Elevar los desvíos al cuadrado, promediarlos, y luego calcular la raíz cuadrada, da un resultado ligeramente diferente al simple promedio de los desvíos de los registros con respecto a la media. Aun así, el resultado de este procedimiento tiene ventajas técnicas que superan la ligera desventaja de dar sólo una descripción aproximada de la variación promedio con respecto a la media (*véase* nota al pie N° 2).

[2] El alumno seguramente se estará preguntando por qué los estadísticos no trabajan sólo con los desvíos, simplemente haciendo que todos los desvíos sean positivos, y utilizando sus promedios. En realidad, en el pasado, ese era el procedimiento. El promedio de los desvíos (tratando a todos los desvíos como positivos) se denomina **desvío promedio** o **desvío medio**. En efecto, algunos psicólogos han hecho resurgir esta cuestión observando algunas ventajas sutiles del desvío promedio (Catanzaro & Taylor, 1996). Sin embargo, a pesar de su simplicidad conceptual y de cálculo, el desvío promedio no funciona muy bien como parte de procedimientos estadísticos más complejos, debido a que resulta difícil realizar manipulaciones algebraicas con una fórmula que ignora los signos de algunos de sus números.

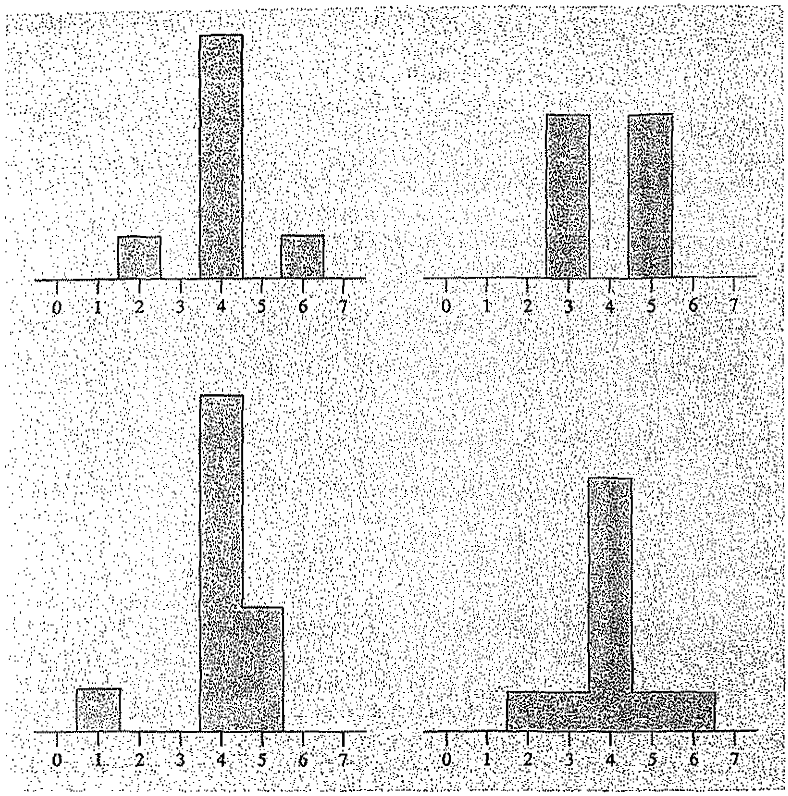


Figura 2-11. Representación gráfica de varias posibles distribuciones de un ejemplo ficticio de composición familiar en el que la media es 4 y el desvío estándar es 1.

Fórmulas para obtener la varianza y el desvío estándar

Hemos visto que la varianza es el desvío cuadrático promedio con respecto a la media. Se calcula con la fórmula:

$$SD^2 = \frac{\Sigma(X-M)^2}{N} \tag{2-2}$$

SD^2 es el símbolo de varianza. (Más adelante aprenderemos sus otros símbolos, S^2 y σ^2, la letra griega "sigma" minúscula al cuadrado. Los diferentes símbolos corresponden a diferentes circunstancias en las que se utiliza la varianza y, en algunos casos, incluso a cálculos ligeramente diferentes).

SD es la abreviatura de **desvío estándar**; recalca que la varianza es el desvío estándar elevado al cuadrado. La parte superior de la fórmula describe la suma de los desvíos cuadráticos.

X se refiere a cada observación en la distribución. M es la media. Por lo tanto, $X - M$ es la observación menos la media, es decir, el desvío. El índice sobrescrito 2 indica que se debe elevar el desvío al cuadrado. Finalmente, el signo de suma (Σ) indica que se deben sumar todos los desvíos cuadráticos.

La suma de los cuadrados es un cálculo importante en muchos procedimientos estadísticos; por lo tanto, tiene su propio símbolo, *SS*. Por esta razón, algunas veces la fórmula de la varianza se escribe utilizando este símbolo en el numerador, en lugar de $\Sigma (X - M)^2$:

$$SD^2 = \frac{SS}{N} \tag{2-3}$$

Ya sea que se utilice el símbolo simplificado *SS* o la descripción completa de la suma de cuadrados, la parte inferior de la fórmula es simplemente N, la cantidad de observaciones. Es decir, la fórmula indica dividir la suma de cuadrados por la cantidad de desvíos cuadráticos (la cantidad de observaciones en la distribución).

El desvío estándar es la raíz cuadrada de la varianza. De modo tal que si se conoce la varianza, la fórmula es simplemente:

$$SD = \sqrt{SD^2} \tag{2-4}$$

La fórmula del desvío estándar comenzando desde el principio es la raíz cuadrada del cálculo de la varianza:

$$SD = \sqrt{\frac{\Sigma(X - M)^2}{N}} \tag{2-5}$$

ó

$$SD = \sqrt{\frac{SS}{N}} \tag{2-6}$$

Ejemplo de cálculo de la varianza y el desvío estándar

La tabla 2-2 ilustra el cálculo de varianza y desvío estándar del ejemplo de las sesiones de terapia. (En la tabla se presume que ya hemos calculado que la media es de 6 sesiones). Generalmente es más sencillo realizar los cálculos utilizando una calculadora, especialmente una que incluya la función para calcular la raíz cuadrada.

Otro ejemplo de cálculo de la varianza y el desvío estándar

La tabla 2-3 muestra el procedimiento aritmético para calcular la varianza y el desvío estándar del ejemplo referido a las interacciones sociales. (Con el fin de ahorrar espacio, la tabla muestra sólo las primeras y últimas observaciones).

Sin entrar en detalles, este resultado significa que la cantidad de interacciones sociales de un estudiante en una semana varía de la media un promedio de 11,49 puntos. El resultado también puede representarse con un histograma (figura 2-12).

Finalmente, analicemos el ejemplo referido al estudio de la cantidad de tiempo requerido para leer oraciones ambiguas. La tabla 2-4 muestra el cálculo de estos datos (otra vez, sólo con las

Tabla 2-2.
Cálculo de varianza y desvío estándar del ejemplo referido a la cantidad de sesiones de terapia.

Observación (cantidad de sesiones)	–	Media (cantidad media de sesiones)	=	Desvío	Desvío cuadrático
7		6		1	1
8		6		2	4
8		6		2	4
7		6		1	1
3		6		–3	9
1		6		–5	25
6		6		0	0
9		6		3	9
3		6		–3	9
8		6		2	4
				Σ: 0	66

$$\text{Varianza} = SD^2 = \frac{\Sigma(X-M)^2}{N} = \frac{SS}{N} = \frac{66}{10} = 6{,}6$$

$$\text{Desvío estándar} = SD = \sqrt{SD^2} = \sqrt{6{,}6} = 2{,}57$$

primeras y últimas observaciones para ahorrar espacio). En términos aproximados, el tiempo promedio que tarda un participante en leer una oración ambigua varía 0,142 segundos de la media de 2,755 segundos. La figura 2-13 representa los datos mencionados.

Tabla 2-3.
Cálculo de la varianza y el desvío estándar de la cantidad de interacciones sociales vividas por 94 estudiantes universitarios durante una semana.

Cantidad de interacciones	–	media de interacciones	=	Desvío	Desvío cuadrático
48		17,40		30,60	936,36
15		17,40		–2,40	5,76
33		17,40		15,60	243,36
3		17,40		–14,40	207,36
21		17,40		3,60	12,96
.		.		.	.
.		.		.	.
.		.		.	.
35		17,40		17,60	309,76
9		17,40		–8,40	70,56
30		17,40		12,60	158,76
8		17,40		–9,40	88,36
26		17,40		8,60	73,96
				Σ: 0,00	12,406,44

$$\text{Varianza} = SD^2 = \frac{\Sigma(X-M)^2}{N} = \frac{12.406{,}44}{94} = 131{,}98$$

$$\text{Desvío estándar} = \sqrt{SD^2} = \sqrt{131{,}98} = 11{,}49$$

Fuente: McLaughlin-Volpe et al. (1998).

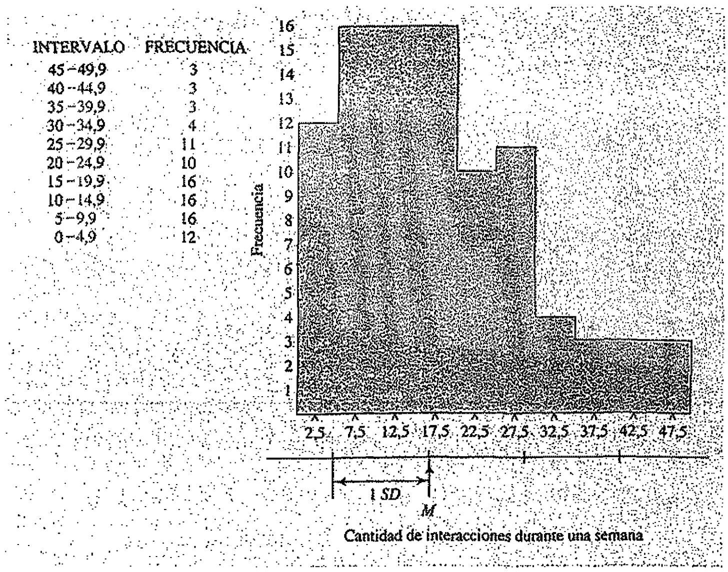


Figura 2-12. Descripción gráfica del desvío estándar como la distancia a lo largo de la base de un histograma, utilizando el ejemplo referido a la cantidad de interacciones sociales vividas en una semana. (**Fuente:** McLaughlin-Volpe et al., 1998).

Tabla 2-4.
Cálculo de la varianza y el desvío estándar en un estudio ficticio del tiempo de lectura de oraciones ambiguas.

Valor (tiempo de lectura)	–	Media (tiempo de lectura)	=	**Desvío**	**Desvío cuadrático**
2,72		2,755		– 0,035	0,0012
2,84		2,755		0,085	0,0072
2,63		2,755		– 0,125	0,0156
2,51		2,755		– 0,245	0,0600
2,54		2,755		– 0,215	0,0462
2,98		2,755		0,225	0,0506
.		.		.	.
.		.		.	.
.		.		.	.
2,52		2,755		– 0,235	0,0552
2,66		2,755		– 0,095	0,0090
2,74		2,755		– 0,015	0,0002
2,73		2,755		– 0,025	0,0006
2,88		2,755		0,125	0,0156
2,85		2,755		0,095	0,0090
				Σ: 0,000	2,0330

$$\text{Varianza} = SD^2 = \frac{\Sigma(X-M)^2}{N} = \frac{SS}{N} = \frac{2{,}033}{100} = 0{,}0203$$

$$\text{Desvío estándar} = SD = \sqrt{SD^2} = \sqrt{0{,}0203} = 0{,}142$$

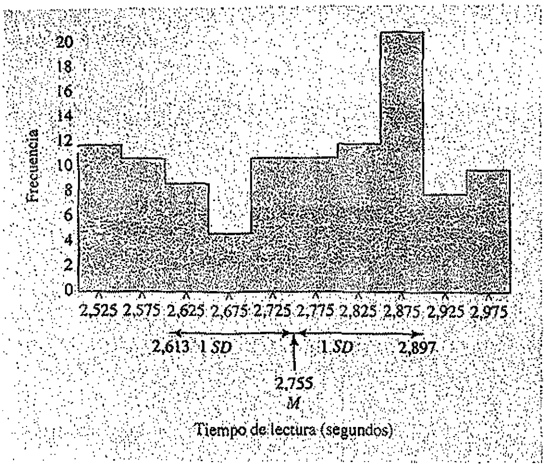


Figura 2-13. Descripciones gráficas del *desvío estándar como la distancia a lo largo de la base de un histograma,* utilizando el ejemplo referido al tiempo necesario (en segundos) para leer oraciones ambiguas (datos ficticios).

Fórmulas de cálculo de la varianza y el desvío estándar

En investigaciones reales, los psicólogos frecuentemente deben calcular la varianza y el desvío estándar de distribuciones que involucran una gran cantidad de casos, y los cálculos generalmente incluyen decimales o grandes números. Debido a ello, el proceso puede llevar demasiado tiempo, aun utilizando una calculadora. Para resolver este problema, se desarrollaron ciertos métodos que simplifican los cálculos. Una fórmula simplificada con el fin arriba mencionado se denomina fórmula de cálculo En el apéndice *de éste capítulo presentamos* las **fórmulas de cálculo** de la varianza y el desvío estándar.

Sin embargo, en la actualidad, las fórmulas de cálculo son interesantes principalmente desde el punto de vista histórico. Los investigadores las utilizan sólo cuando no disponen de computadoras para realizar los cálculos. De hecho, incluso muchas calculadoras están preparadas de tal modo que sólo es necesario ingresar los datos y presionar una o dos teclas para obtener la varianza y el desvío estándar.

En este libro presentamos las fórmulas de cálculo en los apéndices de varios capítulos, en caso de que algún alumno esté realizando un proyecto de investigación que incluya demasiados números y no disponga de una computadora. Sin embargo, no consideramos que las fórmulas de cálculo sean útiles para aprender estadística. Más bien, tienden a oscurecer el significado de los cálculos. Al realizar los ejercicios, conviene utilizar las fórmulas que presentamos en cada capítulo, ya que esas fórmulas han sido diseñadas para ayudar a profundizar la comprensión del significado de los cálculos. Las fórmulas que presentamos en los capítulos se denominan **fórmulas de definición.**

El propósito de este libro es ayudar a comprender los procedimientos estadísticos, y no convertir al alumno en una computadora, haciéndole memorizar y aplicar fórmulas de cálculo que rara vez volverá a utilizar, si es que alguna vez lo hace. (De todos modos, para simplificar los cálculos reales, nuestros ejercicios generalmente utilizan pequeños grupos de números enteros.

Para los alumnos que disponen de una computadora, la *Guía de estudio y cuaderno de ejercicios para computadora [Study Guide and Computer Workbook],* que acompaña este libro, incluye material preparado especialmente para que adquieran experiencia realizando estadísticas del modo en que normalmente lo harían los psicólogos, trabajando con programas estadísticos estandarizados en una computadora).

La varianza como la suma de cuadrados dividida por N - 1

Un aspecto que debemos tener en cuenta es que los psicólogos a menudo utilizan una fórmula de la varianza ligeramente diferente a la que hemos visto, ya sea para definirla o calcularla. Nosotros hemos definido la varianza como el promedio de los desvíos cuadráticos, es decir, como la suma de cuadrados dividida por la cantidad de observaciones, SS/N. Sin embargo, en el capítulo 9 veremos que en muchos casos es correcto definir la varianza como la suma de cuadrados dividida por **la cantidad de observaciones menos 1**; en esos casos, la varianza es $SS/(N-1)$.

La varianza (o su raíz cuadrada, el desvío estándar) que aparece en las publicaciones científicas, con frecuencia se calcula utilizando $SS/(N-1)$. Incluso cuando las calculadoras o computadoras calculan automáticamente la varianza o el desvío estándar, a veces lo hacen con ese mismo método. Pero no debemos preocuparnos por esto ahora. El método que estamos aprendiendo en este capítulo, la varianza = SS/N, es completamente correcto para los fines para los que lo hemos estado utilizando (describir la variación de un grupo de registros), para el material tratado en el resto de este capítulo (puntuaciones Z) y para el material que veremos en los capítulos 3 al 8. Mencionamos el otro método aquí, la varianza = $SS/(N-1)$, sólo para evitar cualquier confusión que pueda surgir cuando el alumno lea sobre la varianza o el desvío estándar en otros textos, o si la calculadora o el programa de la computadora que el alumno utiliza arroja un número para el desvío estándar que parece equivocado. Para simplificar las cosas, no trataremos la razón de ser y el uso del método $N-1$ sino hasta que sea necesario, es decir, a partir del capítulo 9.

PUNTUACIONES Z

Hasta aquí hemos aprendido a describir una distribución de observaciones en función de la media y la varianza. En esta sección, aprenderemos cómo describir una observación en particular según el lugar que ocupe dentro del grupo de observaciones en conjunto. Es decir, aprenderemos a describir una observación según la misma se encuentre sobre o debajo del promedio y según a qué distancia hacia abajo o por encima del mismo esté ubicada.

Supongamos que nos informan que alguien llamado Alan tomó 9 sesiones con la psicoterapeuta (la misma a la cual nos hemos referido en este capítulo). Supongamos también que desconocíamos la cantidad de sesiones tomadas por otros pacientes con la misma terapeuta. En ese caso, sería difícil decir si Alan asistió a muchas o pocas sesiones en relación con otros pacientes.

Sin embargo, supongamos que sí sabemos que la media es 6 y el desvío estándar es 2,57. Con esos datos, queda claro que Alan asistió a una cantidad de sesiones superior al promedio. También podemos ver que la cantidad de sesiones en las que Alan se excedió del promedio (3 sesiones más) era un poco más alta que la cantidad de sesiones en que los pacientes de la terapeuta generalmente varían con respecto al promedio. La figura 2-14 muestra el caso gráficamente.

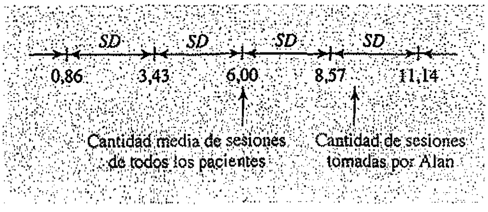


Figura 2-14. Relación entre la cantidad de sesiones tomadas por un paciente llamado Alan y la distribución general de las cantidades de sesiones tomadas por todos los pacientes de una terapeuta en particular (datos ficticios).

Cuadro 2-1. La verdadera alegría (sí, alegría) que provoca el análisis estadístico.

¿Estamos aprendiendo estadística porque nos gusta, verdad? ¿No? O tal vez sí, después de todo. Porque si nos convertimos en psicólogos, en algún momento plantearemos una hipótesis, recolectaremos datos y los analizaremos. (Aun cuando planifiquemos nuestras carreras como psicoterapeutas, es posible que con el tiempo deseemos someter a prueba alguna idea sobre la naturaleza de nuestros pacientes y sus dificultades). Esa hipótesis, nuestra propia idea original, y los datos recolectados para probarla serán muy importantes para nosotros. Incluso es posible que nuestro corazón llegue a latir emocionado al analizar las estadísticas.

Veamos algunos comentarios de psicólogos sociales que entrevistamos para nuestro libro *El corazón de la psicología social [The Heart of Social Psychology]* (Aron & Aron, 1989). Deborah Richardson, quien estudia relaciones interpersonales, nos confió que para ella lo mejor de ser psicóloga social es observar los resultados estadísticos de los análisis computarizados:

> Es como armar un rompecabezas... Es una experiencia muy emocionante y positiva para mí. A menudo tengo periodos de euforia. Incluso cuando la información no arroja el resultado que yo espero.... [existe una] respuesta psicológica ... Es emocionante ver cómo van surgiendo los datos, ¿es realmente como yo pensé que sería? y luego, pensar en las alternativas.

Harry Reis, ex editor de la sección de Proceso Grupal y Relaciones Interpersonales de la *Revista Científica de Psicología Social y de la Personalidad [Journal of Personality and Social Psychology]*, ve su profesión del mismo modo:

> La mejor recompensa es, por mucho, cuando uno obtiene nueva información y comienza a analizarla y comienzan a surgir cosas que, al principio, y hasta cierto punto, son la confirmación de lo que nos indujo a realizar el estudio, pero después también surgen otras cosas... "¿Por qué ocurre esto?". Uno intenta encontrarle sentido. Son el tipo de ideas que surgen de los datos... Adoro analizarlos.

Bibb Latane, un eminente psicólogo conocido, entre otras cosas, por su trabajo sobre la razón por la cual las personas no siempre intervienen para ayudar a otros que se encuentran en problemas, cuenta cómo esperaba ansiosamente:

> Los primeros resplandores de lo que surgió ... [y] poder utilizarlos para formular lo que debería ser la pregunta siguiente... Uno necesita utilizar todo lo que tiene [...] cada pizca de experiencia e intuición. Es cuando se obtiene el mayor efecto, es lo menos parecido a la rutina. Uno está frente a la realidad, cara a cara con la esencia de lo que está desarrollando, en el momento de la verdad.

Bill Graziano, cuyo trabajo integra la psicología social y de desarrollo, llama "gran diversión, sólo gran diversión" al análisis de su información. Y del mismo modo, Margaret Clark, quien estudia la emoción y la cognición, declara que "lo más divertido es obtener datos y analizarlos".

Queda claro entonces que la estadística al servicio de nuestras propias ideas creativas puede ser realmente un placer.

¿Qué es una puntuación Z?

Una **puntación Z** es la transformación de una observación que describe mejor el lugar que esa observación ocupa en la distribución. Específicamente, una puntuación Z indica a qué cantidad de desvíos estándar por encima de la media se encuentra dicha observación (si es positivo), o bien por debajo de la media (si es negativo). El desvío estándar se transforma así en una especie de patrón, una unidad de medida propiamente dicha. En el ejemplo de la psicoterapia, Alan, que asistió a 9 sesiones, tiene una puntación Z de +1,17. Es decir que Alan está 1,17 desvíos estándar por encima de la media (un poco más de 1 desvío estándar de 2,57 sesiones por encima de la media). Otra paciente, Sarah, asistió a 6 sesiones con la terapeuta, y presenta una puntuación Z de 0, ya que el valor que le corresponde es coincidente con la media. Es decir, su valor observado es de 0 desvíos estándar por encima o por debajo de la media. ¿Qué pasaría con un paciente que asistiera sólo a una sesión? Ese paciente habría asistido a 5 sesiones menos que el promedio, casi 2 desvíos estándar por debajo de la media (una puntuación Z de −1,95). En función de la cantidad de sesiones, el paciente estaría por debajo del promedio el doble de veces de lo que varían típicamente con respecto al promedio los pacientes de la terapeuta.

Puntaciones Z utilizadas como escala

La figura 2-15 muestra, utilizando el ejemplo de las sesiones de terapia, una escala de puntuaciones Z dispuesta en comparación con una escala de **puntuación bruta**. (Una puntación bruta es un valor observado antes de ser convertido en una puntuación Z). Las dos escalas se asemejan a una regla con pulgadas de un lado y centímetros del otro, o a un termómetro con la escala Fahrenheit de un lado y la de Celsius del otro.

Otros ejemplos

En la práctica, las puntuaciones Z tienen muchos usos. También son parte importante de muchos de los procedimientos estadísticos que aprenderemos en lo que resta del libro. Es importante familiarizarse con ellos.

Analicemos otro ejemplo. Supongamos que un psicólogo especializado en el desarrollo observó a un niño de tres años, llamado Peter, en una situación estándar de laboratorio, mientras jugaba con otros niños de su edad. Durante la observación, el psicólogo controló la cantidad de veces que Peter hablaba con los otros niños. El resultado, luego de varias observaciones, fue que Peter habló con los otros niños aproximadamente 8 veces por hora de juego. Sin ningún patrón de comparación, sería difícil sacar alguna conclusión a partir de esta información. Supongamos, sin embargo, que se sabía, por investigaciones previas, que en similares condiciones la cantidad media de veces que los niños hablan por hora de juego es 12, con un desvío estándar de 4. Con esa información, ahora podemos ver que Peter habló con menos frecuencia que otros niños en general, pero no con una frecuencia extremadamente menor. Peter tendría una puntuación Z de −1 (si $M = 12$ y $SD = 4$, una observación de 8 está 1 *SD* por debajo de la *M*). Supongamos que observamos conversar a Ian con otros niños 20 veces en una hora. Quedaría claro que Ian es inusualmente locuaz, con una puntuación Z de + 2. Ian no sólo hablaría más que el promedio, sino dos veces más de lo que los niños tienden a desviarse del promedio. (*Véase* figura 2-16).

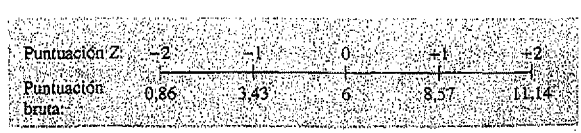


Figura 2-15. Escalas de puntuaciones Z y puntuación bruta en el caso de las sesiones de psicoterapia.

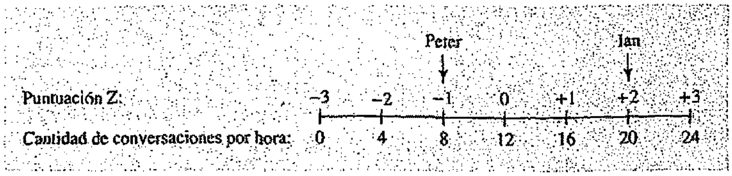


Figura 2-16. Cantidad de veces por hora que dos niños conversan, expresada en puntuaciones brutas y puntuaciones Z (datos ficticios).

Puntuaciones Z como patrones de comparación generalizados

Otra ventaja de las puntuaciones Z es que, convirtiendo las observaciones de variables completamente diferentes en puntuaciones Z, podemos compararlas entre sí. Con las puntuaciones Z, la media es siempre 0 y el desvío estándar es siempre 1. Supongamos que los mismos niños de nuestro ejemplo fueran medidos en una prueba de capacidad lingüística. Podríamos comparar directamente las puntuaciones Z correspondientes a esa capacidad con las puntuaciones Z correspondientes al nivel de conversación con otros niños. Supongamos que Peter obtuvo una puntuación bruta de 100 en la prueba lingüística. Si la media en la prueba era 82 y el desvío estándar era 6, entonces Peter tiene una capacidad lingüística claramente superior al promedio, con una puntuación Z de + 3. Es improbable que el grado de conversación con otros niños, menor al usual, demostrado por Peter, se deba a una capacidad lingüística menor a la usual (*véase* figura 2-17).

El ejemplo nos muestra que, utilizando puntuaciones Z, podemos comparar directamente los resultados de observaciones psicológicas del grado de conversación con los resultados de una prueba de capacidad lingüística. ¡Esto es casi tan fabuloso como poder comparar *manzanas con* naranjas! Convertir un número en una puntuación Z se asemeja a convertir los términos y unidades de medida desconocidos en otros que todos podamos entender; es como convertir codos y pulgadas, por ejemplo, en centímetros. Es realmente una herramienta muy valiosa.

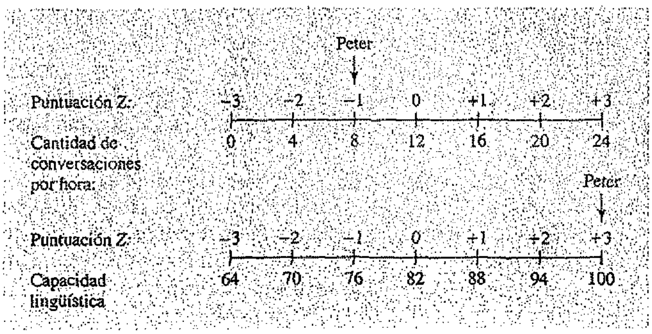


Figura 2-17. Escalas de puntuación Z y puntaciones brutas para la cantidad de conversaciones por hora y la capacidad lingüística, con las observaciones del primer niño en cada una de ellas (datos ficticios).

Fórmula para convertir una puntuación bruta en una puntuación Z

Una observación directa se denomina puntuación bruta. Como hemos observado, una puntuación Z indica la cantidad de desvíos estándar por encima de la media (o, si es negativo, por debajo) en que se encuentra la puntuación bruta. Para calcular una puntuación Z, se resta la media a la puntuación bruta, obteniendo el desvío. Luego se divide el desvío por el desvío estándar. En símbolos, la fórmula es la siguiente:

$$Z = \frac{X - M}{SD} \tag{2-7}$$

Por ejemplo, si aplicamos la fórmula al ejemplo del niño con un registro de 100 en la prueba de capacidad lingüística, la fórmula sería la siguiente:

$$Z = \frac{X - M}{SD} = \frac{100 - 82}{6} = \frac{18}{6} = 3$$

Fórmula para la conversión de una puntuación Z en una puntuación bruta

Para volver a convertir una puntuación Z en una puntuación bruta se invierte el proceso: se multiplica la puntuación Z por el desvío estándar y luego se le suma la media. La fórmula es la siguiente:

$$X = (Z)(SD) + M \tag{2-8}$$

Por ejemplo, si un niño presenta una puntuación Z de –1,5 en la prueba de capacidad lingüística, quiere decir que se encuentra 1,5 desvíos estándar por debajo de la media. Dado que el desvío estándar en este caso es de 6 puntos brutos, el niño está 9 puntos brutos por debajo de la media. La media es 82. Por lo tanto, 9 puntos por debajo de ella es 73. Utilizando la fórmula, tendríamos:

$$X = (Z)(SD) + M = (-1{,}5)(6) + 82 = -9 + 82 = 73$$

Pasos para convertir una puntuación bruta en una puntuación Z

Para convertir una puntuación bruta en una puntuación Z:

1. Calcular el desvío: restar la media a la puntuación bruta.
2. Calcular la puntuación Z: dividir el desvío por el desvío estándar.

Pasos para convertir una puntuación Z en una puntuación bruta

Para convertir una puntuación Z en una puntuación bruta:

1. Calcular el desvío: multiplicar la puntuación Z por el desvío estándar.
2. Calcular la puntuación bruta: sumar la media al desvío.

Ejemplos de cálculo de una puntuación Z a partir de una puntuación bruta y viceversa

Analicemos el primer ejemplo que utilizamos en el capítulo 1, es decir, las puntuaciones de estrés de 151 alumnos de estadística (Aron et al., 1995). La media de esa distribución era 6,46, y el desvío estándar era de 2,30 ($SS = 797{,}5$; $SD^2 = 797{,}5/151 = 5{,}28$; $SD = 5{,}28 = 2{,}30$).

La figura 2-18 muestra la relación entre las escalas de puntuaciones brutas y de puntuaciones Z. Si la puntuación bruta del estrés de un alumno era 9, ese alumno se encuentra claramente por encima de la media. Específicamente, al utilizar la fórmula obtendríamos:

$$Z = \frac{X - M}{SD} = \frac{9 - 6,46}{2,3} = \frac{2,54}{2,3} = 1,10 \qquad (2\text{-}7)$$

En comparación, otro alumno presenta una puntuación Z de −2,37, un nivel de estrés claramente menor a la media. Utilizando la fórmula, la puntuación bruta exacta se calcula de la siguiente forma:

$$X = (Z)(SD) + M = (-2,37)(2,3) + 6,46 = -5,45 + 6,46 = 1,0$$

Analicemos algunos ejemplos tomados del estudio sobre la cantidad de interacciones sociales vividas por estudiantes en una semana (McLaughlin-Volpe et al., 1998). Recordemos que la media era 17,4 y el desvío estándar 11,49. Un alumno que tuvo 17 interacciones en una semana presenta un desvío de − 0,4 (es decir 17 − 17,4 = − 0,4). Por lo tanto, la puntuación Z es − 0,03 (es decir −0,4/11,49 = −0,03). Esta cantidad de interacciones se encuentra apenas por debajo de la media. De modo similar, un estudiante que vivió 36 interacciones sociales en una semana presenta un desvío de 18,6 (es decir, 36 − 17,4 = 18,6). La puntuación Z es 1,62 (es decir 18,6/11,49 = 1,62). Este alumno se encuentra a 1,62 desvíos estándar por sobre la media en lo que se refiere a interacciones sociales vividas en el plazo de una semana.

Para hacerlo del modo inverso, supongamos que supiéramos que la puntuación Z de un alumno fue 0,57. La puntuación bruta de ese alumno (referido a la cantidad de interacciones sociales) sería igual a la puntuación Z por el desvío estándar más la media: (0,57 x 11,49) + 17,4 = 23,95 (redondeando, esta persona tuvo 24 interacciones sociales). La figura 2-19 muestra estas relaciones.

Finalmente, analicemos el ejemplo referido al tiempo de lectura de oraciones. Recordemos que en ese estudio calculamos que la media del tiempo de lectura de 100 estudiantes era 2,755 segundos y el desvío estándar de 0,143 segundos. Un alumno con un tiempo de lectura de 2,88 segundos presenta una puntuación Z calculada de la siguiente forma:

$$Z = \frac{X - M}{SD} = \frac{2,88 - 2,755}{0,143} = \frac{0,125}{0,143} = 0,87$$

Un alumno con una puntuación Z de 2,0 presenta un tiempo de lectura calculado de la siguiente forma:

$$X = (Z)(SD) + M = (2,0)(0,143) + 2,755 = 0,286 + 2,755 = 3,041$$

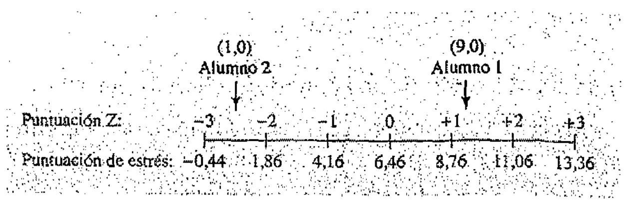


Figura 2-18. Escalas de puntuaciones brutas y puntuaciones Z referidas a los valores de estrés de 151 alumnos de estadística (**Fuente**: Aron, Paris, & Aron, 1995), con las observaciones de dos alumnos tomados como muestra.

El tiempo de lectura de un alumno con una puntuación Z de $-1,1$ es el siguiente:

$$X = (Z)(SD) + M = (-1,1)(0,143) + 2,755 = -0,157 + 2,755 = 2,598$$

La figura 2-20 ilustra estas relaciones.

Algunas características de las puntuaciones Z

La media de una distribución de puntuaciones Z siempre es exactamente 0, debido a que la conversión de observaciones a puntuaciones Z involucra restar la media de cada puntuación bruta. Para decirlo de otro modo, la suma de las puntuaciones Z positivas de una distribución siempre debe ser igual a la suma de las puntuaciones Z negativas de dicha distribución.

El desvío estándar de una distribución de puntuaciones Z siempre es exactamente 1, debido a que la conversión de observaciones a puntuaciones Z involucra dividir cada desvío por el desvío estándar. Además, dado que el desvío estándar es 1, la varianza, es decir el desvío estándar elevado al cuadrado, también es siempre 1. La tabla 2-5 indica las puntuaciones Z correspondientes al estudio de las sesiones de terapia, junto con los cálculos de la media y el desvío estándar de esas puntuaciones Z. Este ejemplo demuestra que, tratándose de puntuaciones Z, la media es 0 y el desvío estándar (y varianza) es 1.

Las puntuaciones Z se denominan a veces **puntuación estándar**, debido a que presentan valores estándares para la media y el desvío estándar. También se debe a que, como vimos anteriormente, las puntuaciones Z brindan una especie de escala de medición estándar para cualquier variable. (Sin embargo, a veces el término **puntuación estándar** se utiliza sólo cuando las puntuaciones Z se refieren a una distribución que es una curva normal. Como veremos más adelante, en el capítulo 5, las puntuaciones Z son aun más útiles cuando la distribución es una curva normal).

Además, a veces se calculan puntuaciones similares a puntuaciones Z cuya media es un número distinto de 0 y cuyo desvío estándar es un número distinto de 1. Por ejemplo, en algunas pruebas utilizadas por psicólogos clínicos, se crean escalas especiales en las que la media es 50 y el desvío estándar es 10. Por lo tanto, un registro de 65, en esta escala, sería igual a una puntuación Z de 1,5.

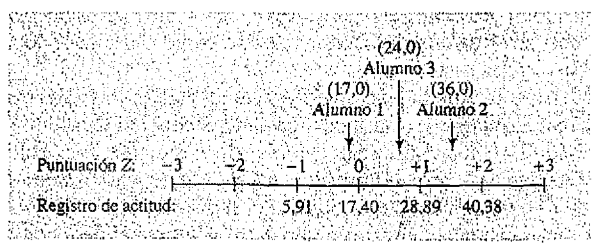


Figura 2-19. Escalas de puntuaciones brutas y puntuaciones Z referidas a la cantidad de interacciones sociales vividas por 94 alumnos durante una semana, con las observaciones de tres alumnos tomados como muestra. (**Fuente:** McLaughlin-Volpe et al., 1998).

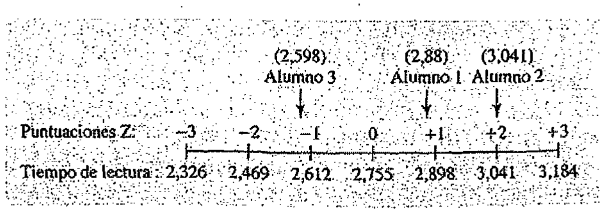


Figura 2-20. Escalas de puntuaciones brutas y puntuaciones Z referidas al tiempo de lectura de oraciones ambiguas por parte de 100 alumnos (datos ficticios), con los registros de tres casos tomados como muestra.

Tabla 2-5.
Cálculo de la media y el desvío estándar de puntuaciones Z en el ejemplo de la cantidad de sesiones de terapia.

Cantidad de sesiones (Puntuación bruta)	Puntuación Z de la cantidad de sesiones	Media punto Z	Desvío de la puntuación Z	Desvío cuadrático de la puntuación Z
7	0,39	0	0,39	0,15
8	0,78	0	0,78	0,61
8	0,78	0	0,78	0,61
7	0,39	0	0,39	0,15
3	-1,17	0	-1,17	1,37
1	-1,95	0	-1,95	3,80
6	0,00	0	0,00	0,00
9	1,17	0	1,17	1,37
3	-1,17	0	-1,17	1,37
8	0,78	0	0,78	0,61
	Σ: 0,00			10,04[a]

$$M = \frac{\Sigma X}{N} = \frac{0}{10} = 0$$

$$SD^2 = \frac{\Sigma(X-M)^2}{N} = \frac{SS}{N} = \frac{10}{10} = 1$$

$$SD = \sqrt{1} = 1$$

[a] Si no hubiera errores de redondeo, el resultado sería igual a 10.

CONTROVERSIAS Y LIMITACIONES: LA TIRANÍA DE LA MEDIA

Aun cuando el uso de la estadística en psicología es tan generalizado que pareciera ser la única herramienta o el único lenguaje de esta disciplina, siempre ha existido una corriente en desacuerdo con el método puramente numérico. Es nuestra intención informar al lector, a lo largo del libro, sobre las controversias que existen en el campo de la psicología con respecto a la estadística. Consideramos que un buen tema para comenzar a hacerlo es precisamente el debate referido al abuso de las estadísticas.

El "padre de la psicología", Wihelm Wundt, pensaba que los experimentos y las estadísticas debían limitarse a temas tales como la percepción y la memoria, una opinión que rara vez se menciona. El método apropiado para las otras áreas de la psicología era el análisis y la interpretación del significado, procedimientos que prescinden de los números (McLeod, 1996).

El conductismo se describe con frecuencia como la escuela de psicología históricamente más dedicada a mantener este campo dentro de un ámbito estrictamente científico. El conductismo se inició alrededor del año 1913, con el rechazo por el estudio de los estados interiores del individuo debido a la imposibilidad de observarlos objetivamente. Pero el más ardiente portavoz del conductismo, B. F. Skinner, se oponía rotundamente a la estadística. Skinner llegó incluso a decir: "Preferiría ver a un graduado en psicología asistir a un curso de físico-química que de estadística. E incluiría (presumiblemente antes que la estadística) otras ciencias, incluso poesía, música y arte" (Evans, 1976, p. 93).

¿Por qué Skinner se oponía tan rotundamente a la estadística? Él sostenía que observar el comportamiento es la mejor forma de comprenderlo, y se refería a la observación de casos individuales.

Hacía notar constantemente los datos que se perdían por promediar los resultados de varios casos. Por ejemplo, Skinner (1956) mencionaba el ejemplo de tres ratones que comían en exceso: uno naturalmente obeso, otro envenenado con oro y otro cuyo hipotálamo había sido alterado. Cada uno presentaba una curva de aprendizaje diferente (patrón de velocidad de aprendizaje) en relación con la destreza necesaria para presionar una barra y alcanzar el alimento; esto revelaba muchos aspectos acerca de los hábitos alimenticios ocasionados por cada una de las distintas enfermedades. Si se hubieran sumado o unificado estadísticamente las curvas de aprendizaje, el resultado no hubiera representado los hábitos alimenticios reales de ningún ratón real. Según el mismo Skinner, "estas tres curvas individuales contienen más información de la que podría haber sido generada por medidas que requirieran un tratamiento estadístico; sin embargo, las mismas serán analizadas con desconfianza por muchos psicólogos porque representan casos individuales". (p. 232)

Diferente fue el pedido de precaución emitido por la psicología humanística, cuyos comienzos datan de la década de 1950 como "tercera fuerza" en contraposición al conductismo y a la principal alternativa del momento, el psicoanálisis freudiano. El tema central de la psicología humanística establecía que la conciencia humana debía ser estudiada íntegramente, como un todo, exactamente como es experimentada por el individuo. No es posible explicar totalmente la experiencia humana reduciéndola a números (así como tampoco es posible explicarla reduciéndola a palabras). La experiencia de un individuo es compleja y única.

En el área de la psicología clínica y del estudio de la personalidad, a menudo se han levantado voces para argumentar que puede aprenderse mucho más sobre aquello que es realmente importante en psicología a partir del análisis profundo de una persona, que respecto de promedios entre varias de ellas. Es decir, el método **idiográfico** contra el **nomotético**, para utilizar los términos que Gordon Allport tomó de Wilhelm Windelband (*véase* Hilgard, 1987). Y la base filosófica del análisis profundo de los individuos puede encontrarse en la fenomenología, que nació en Europa después de la Primera Guerra Mundial (*véase* Husserl, 1970).

La fenomenología es una posición filosófica opuesta al positivismo lógico. El positivismo lógico sostiene que existe una realidad objetiva a ser conocida. Es la posición filosófica que sustenta tradicionalmente los esfuerzos científicos. Se considera que la ciencia puede descubrir esa realidad objetiva o verdadera dado que utiliza experimentos que cualquiera puede observar o repetir para obtener los mismos resultados. Los fenomenólogos sostienen, sin embargo, que incluso estas reiteradas observaciones son en realidad hechos particulares realizados en forma consciente. Uno no puede saber si lo que entiende por "verde" o "la rata presionó la barra siete veces" es lo que cualquier otro entiende por esas mismas palabras. Según los fenomenólogos, no existe una realidad objetiva de la cual todos podamos estar seguros.

En la actualidad, el principal desafío para la estadística proviene del fuerte renacimiento del interés en los métodos "cualitativos" de investigación. Ha habido una creciente preocupación entre algunos psicólogos con respecto a que, luego de cien años de investigación estadística cuantitativa, la psicología ha producido lo que ellos consideran conocimientos de muy poca utilidad social (Jessor, 1996). Esperan que, analizando cuidadosamente como un todo a unos pocos seres humanos en su contexto se puedan obtener mejores resultados

Highlen y Finley (1996) describen cinco posibles posiciones filosóficas que acompañan la investigación cualitativa. La primera adopta el positivismo lógico y busca una realidad objetiva a través de métodos cualitativos. También existe el **pospositivismo**, que sostiene la existencia de una realidad verdadera pero que nunca conoceremos completamente. No obstante, esforzándonos podemos acercarnos a ella. La visión del **constructivismo** subraya la existencia de múltiples realidades. Cada uno de nosotros construye un significado a partir de la experiencia, y la psicología debería intentar comprender algunos de esos significados. La **visión crítica** también niega cualquier realidad objetiva. Sostiene que toda ciencia sirve al propósito de alguien, y el propósito correcto es la liberación de los más débiles a través de, por ejemplo, el feminismo o el neomarxismo. Finalmente, la visión **postestructural** persigue el objetivo de desafiar toda realidad socialmente establecida, la cual es considerada el producto de quienquiera que detente el poder. Si el alumno aún no ha considerado este tema, aconsejamos averiguar y leer al respecto para comenzar a formar una opinión propia.

Cualquiera sea la posición filosófica subyacente, los métodos cualitativos incluyen análisis de casos, etnografía, fenomenología, interaccionismo simbólico, análisis de sistemas e "investigación de la acción" (Highlen & Finley, 1996). Estos métodos se desarrollaron principalmente en antropología, en donde el conductismo y el positivismo lógico nunca tuvieron la influencia que lograron en la psicología. Los métodos cualitativos usualmente implican largas entrevistas u observaciones de unos pocos individuos; mientras se realizan las entrevistas, el investigador altamente capacitado decide qué aspectos merecen ser recordados, registrados y analizados por medio de otras preguntas y observaciones. Según esta postura, la mente del investigador es la herramienta principal, ya que sólo esa mente puede localizar las relaciones importantes entre las muchas categorías de hechos que surgen de las palabras de quien responde.

Algunos psicólogos (p. ej., Kenney, 1995; McCracken, 1988) sostienen que los métodos cuantitativos y cualitativos pueden y deben complementarse. Primero deberíamos descubrir las categorías importantes a través de un enfoque cualitativo, y luego determinar su incidencia en una población mayor a través de métodos cuantitativos. Este grupo de psicólogos sostiene que, con frecuencia, los investigadores cuantitativos deciden apresuradamente cuáles son las categorías importantes sin explorar primero la experiencia humana con respecto a ellas, a través de entrevistas de preguntas abiertas u observaciones.

También resultan de interés las opiniones muy originales del psiquiatra Carl Jung sobre lo que él llamaba "el *estado de ánimo estadístico*". Tal como lo expresara la analista jungiana Marie Louise von Franz (1979), "tenemos un estado de ánimo estadístico cuando caminamos por una calle y observamos los cientos de rostros inexpresivos y comenzamos a sentirnos disminuidos". Nos sentimos simplemente uno más de la multitud, comunes. O bien, cuando estamos enamorados, sentimos que la otra persona es única y maravillosa; no obstante, cuando nuestro estado de ánimo es estadístico, nos damos cuenta de que la otra persona es común, igual a muchas otras.

Von Franz señala, sin embargo, que si sucediera una catástrofe, cada persona respondería de forma única. En la vida existe al menos tanta irregularidad como regularidad.

> El hecho de que esta mesa no levite sino que permanezca donde está sólo se debe a que los miles y miles y miles de millones de electrones que la forman tienden a comportarse de ese modo estadísticamente. Pero cada electrón por sí mismo podría comportarse de modo diferente. (p. IV-17)

Según Franz, el estado de ánimo estadístico es dañino para el amor y la vida. Para contrarrestarlo, "se necesita un acto de lealtad para con nuestros propios sentimientos" (p. IV-18). Los sentimientos "hacen que la vida, al igual que las relaciones y los actos parezcan únicos y les dan un valor definido" (pp. IV-18–IV-19). En particular, sentir la importancia de nuestras acciones individuales hace menos posibles las inmoralidades, como por ejemplo la guerra y el homicidio. No podemos contar los muertos como si fueran números sino que debemos tratarlos como personas, con emociones y objetivos, como nosotros mismos.

Para resumir, podemos decir que siempre han existido buenas razones para limitar nuestro pensamiento estadístico a su propio territorio, y dejar que nuestro corazón gobierne libremente los otros.

LA MEDIA Y EL DESVÍO ESTÁNDAR SEGÚN SE DESCRIBEN EN PUBLICACIONES CIENTÍFICAS

En las publicaciones científicas normalmente se hace referencia a la media y al desvío estándar. Aunque la varianza y las puntuaciones Z son extremadamente importantes como pasos de procedimientos avanzados que aprenderemos más adelante, rara vez son mencionadas en las publicaciones.

En algunas oportunidades, la media y el desvío estándar son incluidos en el texto de una publicación. Por ejemplo, nuestra psicoterapeuta ficticia podría escribir: "La cantidad media de sesiones tomadas por los últimos 10 pacientes fue 6,0 ($SD = 2{,}57$)."

En las tablas, frecuentemente se hace referencia a la media o al desvío estándar, en especial cuando se involucran varios grupos o cuando los participantes en la investigación son analizados en varias condiciones diferentes. Por ejemplo, Orbach y sus colegas (1997), en un estudio realizado en Israel, compararon un grupo de pacientes suicidas de un hospital para enfermos con problemas mentales (individuos que habían realizado intentos serios de suicidio), pacientes no suicidas de un hospital para enfermos con problemas mentales con diagnosis similares, y un grupo de control (voluntarios de la comunidad). El objetivo del estudio era probar la teoría de que los suicidas tienen mayor tolerancia al dolor físico; que su más alto umbral de dolor hace que para ellos sea más sencillo realizar los dolorosos actos que implica un suicidio. Los investigadores realizaron las pruebas de rutina para medir el umbral de dolor y otras sensaciones, y entregaron varios cuestionarios a los tres grupos. La tabla 2-6, reproducción de la que aparece en su artículo, refleja la media de cada grupo en todas las mediciones.

Tabla 2-6.
Medias y desvíos estándar de medidas de dolor, tendencias suicidas, disociación y medidas emocionales del grupo de estudio.

	Suicidas		Psiquiátricos		Normales	
Medición	*M*	*SD*	*M*	*SD*	*M*	*SD*
Umbral de percepción	38,09	5,31	35,93	3,14	33,03	0,52
Umbral de dolor	45,37	4,81	42,28	3,68	40,01	3,63
Tolerancia al dolor	48,29	2,60	46,68	3,04	46,31	4,22
Máxima tolerancia	2,55	2,43	0,66	1,29	1,14	1,91
Estimación de magnitud	0,54	0,12	0,57	0,08	0,59	0,09
Atracción hacia la vida	2,88	0,98	3,48	0,91	4,11	0,56
Atracción hacia la muerte	3,62	1,04	2,62	0,87	2,60	0,80
Repulsión a la vida	3,02	0,92	2,41	0,76	1,93	0,66
Repulsión a la muerte	1,85	0,82	2,80	1,04	2,64	1,01
Disociación afectiva	2,36	0,67	2,12	0,49	2,01	0,42
Disociación cognitiva	2,10	0,72	1,76	0,37	1,77	0,49
Disociación relacionada con el control	2,07	0,73	1,93	0,57	1,68	0,52
Desesperanza	10,06	5,99	5,55	4,23	4,37	3,70
Depresión	2,86	1,21	2,30	0,90	1,84	0,75
Angustia	2,36	0,90	2,04	0,88	1,95	0,69

Nota: Altos índices de atracción hacia la vida y repulsión a la muerte representan bajas tendencias suicidas; bajos índices de repulsión a la muerte y atracción hacia la vida representan altas tendencias suicidas.
Fuente: Orbach, I. et al. (1997), tab. 1. "Umbral y tolerancia al dolor físico en adolescentes suicidas y no suicidas". *Revista Científica de Psicología de Asesoramiento y Clínica [Journal of Consulting and Clinical Psychology]*, 65, 646–652. Copyright, 1997, por la Asociación Americana de Psicología [American Psychological Association]. Reimpreso con autorización.

Como podemos observar en la tabla, coincidentemente con las predicciones de los investigadores, el grupo suicida presentaba un umbral más alto de dolor que los otros dos grupos y difería de éstos también en varias otras medidas. (Cabe destacar especialmente la gran diferencia entre el grupo suicida y los otros dos grupos en cuanto a la "desesperanza"). Por supuesto, tal como lo indican los desvíos estándar, hay mucha superposición entre los grupos con respecto a estas mediciones. Es decir, aunque teniendo en cuenta el promedio, el grupo suicida presenta un mayor umbral de dolor; existen muchos pacientes suicidas con umbrales de dolor menores a los de los otros grupos, y muchos individuos no suicidas con mayor umbral de dolor.

La tabla 2-7 (tomada de Norcross et al., 1996) presenta un ejemplo particularmente interesante. No muestra desvíos estándar pero sí medias y medianas. Por ejemplo, en 1992, la media de aspirantes a doctorados de asesoramiento psicológico era 120,2, pero la mediana era sólo 110. Esto sugiere que existían ciertos programas con una gran cantidad de aspirantes que tornaban asimétrica la distribución. De hecho, podemos ver en la tabla que en casi todos los casos, y tanto para solicitudes como para inscripciones, las medias son usualmente mayores que las medianas. (Es probable que resulte asombrosa la competitividad que presenta el ingreso a un doctorado en muchas de las áreas de la psicología. Según nuestra experiencia, uno de los factores con bastante influencia en este aspecto, es haber tenido éxito en los cursos sobre estadística).

Tabla 2-7.
Estadística de solicitudes e inscripciones por área y año: Doctorados.

				Solicitudes						Inscripciones	
	N° de programas			*M*			*Mdn*			*M*	*Mdn*
Programa	*1973*[a]	*1979*[a]	*1992*	*1973*[a]	*1979*[a]	*1992*	*1973*[a]	*1979*[a]	*1992*	*1992*	*1992*
Clínica	105	130	225	314.4	252,6	191,1	290	234	168	12,0	8
Cognitiva			47			24,6			22	2,6	2
Comunitaria	4	2	5	90,5		24,4	60		23	3,2	2
Asesoramiento	29	43	62	133,4	90,9	120,2	120	84	110	7,3	6
Desarrollo	56	72	97	54,1	38,9	27,6	41	30	24	2,8	2
Educacional	23	28	30	67,8	39,7	20,0	34	26	12	6,0	4
Experimental y general	118	127	78	56,2	33,2	31,3	42	25	26	4,4	3
Salud			7			40,7			30	4,4	5
Industrial/organizacional	20	25	49	39,9	54,7	66,2	37	48	70	4,9	4
Personalidad	23	15	10	42,5	24,7	12,3	33	17	6	1,0	1
Percepcion/psicofísica			15			8,3			6	1,4	1
Fisiología/biopsicología	40	43	76	33,2	29,3	20,0	29	24	20	3,9	2
Escolar	30	39	56	78,5	54,0	31,3	53	34	32	5,4	5
Social	58	72	59	46,7	30,9	47,1	40	24	37	3,3	3
Otras	47	37	273	61,6	74,1	26,6	27	25	15	3,3	2
Total	566	645	1,089	106,1	85,2	69,4			31	5,6	4

Nota: Los años académicos corresponden a las ediciones de postrado de psicología de 1975-1976, 1981-1982, y 1994 respectivamente.
[a]**Fuente**: Stoup y Benjamin (1982).
Fuente: Norcross, J. C., Hanych, J. M., & Terranova, R. D. (1996), tab. 7. Postgrado de Psicología: 1992-1993. *Psicólogo Americano [American Psychologist]*, 51, 631-643. Copyright 1996, por la Asociación Americana de Psicología [American Psychological Association]. Reimpreso con autorización.

Resumen

La media es un promedio común, es decir, la suma de las observaciones dividida por la cantidad de ellas. Expresado en símbolos, $M = \Sigma X/N$.

Otras formas alternativas menos comunes de descripción de la tendencia central de una distribución son la moda (el valor más común) y la mediana (el valor del registro medio después de ordenar todas las observaciones de menor a mayor).

La variación de un grupo de observaciones puede ser descripta a través de la varianza, es decir, el promedio de los desvíos cuadráticos de cada observación con respecto a la media. Expresado en símbolos: $SD^2 = \Sigma(X - M)^2/N$. La suma de los desvíos cuadráticos también se simboliza como SS. Por lo tanto $SD^2 = SS/N$.

El desvío estándar es la raíz cuadrada de la varianza. Expresado en símbolos: $SD = \sqrt{SD^2}$. Para explicarlo en forma más clara, es aproximadamente el promedio de las diferencias entre las observaciones y la media.

Una puntuación Z indica a cuántos desvíos estándar por encima o por debajo de la media se encuentra una puntuación bruta. Entre otras cosas, las puntuaciones Z sirven para comparar observaciones de variables que tienen diferentes escalas.

Siempre ha habido psicólogos que advirtieron los riesgos que implica el uso de la metodología estadística, ya que en el proceso de resumir los datos en un promedio se pierde información sobre cada caso individual.

Las publicaciones científicas generalmente hacen referencia a la media y al desvío estándar, tanto en el texto como en las tablas. En cambio, rara vez se refieren a la varianza y a las puntuaciones Z.

Términos clave

- Tendencia central.
- Fórmulas de cálculo.
- Fórmulas de definición.
- Desvío.
- Media (M).
- Mediana.
- Moda.
- N.
- Puntuaciones brutas.
- Desvío cuadrático.
- Desvío estándar (SD).
- Puntuaciones estándar.
- Suma de cuadrados (SS).
- Varianza (SD^2).
- Puntuaciones Z.
- Σ.

Ejercicios

Los ejercicios implican la realización de cálculos (con la ayuda de una calculadora). La mayoría de los problemas estadísticos reales se resuelven por computadora, pero aunque exista la posibilidad de utilizar una computadora, es conveniente realizar estos ejercicios manualmente para incorporar el método de trabajo.

Para adquirir práctica en la utilización de una computadora, para resolver problemas estadísticos, se puede utilizar la sección de computación de cada capítulo, publicada en la *Guía de estudio y libro de tareas de computación para el alumno [Student's Study Guide and Computer Workbook]* que acompaña este libro.

Todos los datos de esta sección son ficticios (a menos que se especifique lo contrario).

Las respuestas a los ejercicios de la serie I se encuentran al final del libro.

SERIE I

1. Para cada serie de observaciones determine lo siguiente (muestre los pasos a seguir): a) media, b) mediana, c) *SS* (suma de cuadrados), d) varianza y e) desvío estándar.

Serie A: 32, 28, 24, 28, 28, 31, 35, 29, 26.
Serie B: 6, 1, 4, 2, 3, 4, 6, 6.

2. El 26 de diciembre, en Montreal, la temperatura, medida en 10 oportunidades elegidas al azar y en grados Celsius, fue de -5, -4, -1, -1, 0, -8, -5, -9, -13, y -24. Describa la temperatura típica y la variación de la temperatura a una persona que nunca ha asistido a un curso de estadística. Presente tres maneras diferentes de describir la temperatura típica y dos formas de describir la variación, explicando las diferencias entre ellas y cómo fueron calculadas. (Aprenderá más si intenta escribir primero su propia respuesta, antes de leer la nuestra. Su propia respuesta no necesariamente debe ser tan completa como la respuesta modelo del libro).

3. Se realiza un estudio sobre la cantidad de sueños narrados por 30 personas en psicoterapia, durante un período de dos semanas. En una publicación que describe los resultados, los autores informan: "La cantidad media de sueños fue 6,84 (SD = 3,18)." Explique el significado del enunciado anterior a una persona que nunca ha asistido a un curso de estadística.

4. En una medición de angustia, la media es 79 y el desvío estándar es 12. ¿Cuáles son las puntuaciones Z correspondientes a cada una de las siguientes puntuaciones brutas? a) 81, b) 68, c) 103.

5. En una prueba de inteligencia en particular, la cantidad media de ítems correctos es 231, y el desvío estándar 41. ¿Cuáles son las puntuaciones brutas en esta prueba para personas con CI (Cociente intelectual) de a) 107, b) 83 y c) 100? (El CI es igual a 100, más 16 multiplicado por la puntuación Z. La medida de los CI es 100 y el desvío estándar 16.) (Nota: para resolver este problema, primero calcule la puntuación Z correspondiente a cada CI; luego utilice esa puntuación Z para calcular la puntuación bruta).

6. Seis meses después de divorciarse, cada uno de los ex esposos de una pareja realiza una prueba para medir su adaptación al divorcio. El registro de la esposa es 63 y el del esposo 59. Por lo general, la media para mujeres divorciadas que realizan esta prueba es 60 (SD = 6); la media para hombres divorciados es 55 (SD = 4). ¿Cuál de los dos se ha adaptado mejor al divorcio en relación con otras personas divorciadas del mismo sexo? Explique su respuesta a una persona que nunca ha asistido a un curso de estadística.

SERIE II

1. Defina media, mediana y moda. Cite un caso en el que la mediana sería la medida preferida de la tendencia central.

2. Para cada serie de observaciones determine lo siguiente (muestre los pasos a seguir): a) media, b) mediana, c) *SS* (suma de cuadrados), d) varianza y e) desvío estándar.
Serie A: 2, 2, 0, 5, 1, 4, 1, 3, 0, 0, 1, 4, 4, 0, 1, 4, 3, 4, 2, 1, 0
Serie B: 1.112, 1.245, 1.361, 1.372, 1.472
Serie C: 3,0, 3,4, 2,6, 3,3, 3,5, 3,2

3. Un psicólogo interesado en el comportamiento político midió los pies cuadrados de los escritorios de los despachos de cuatro gobernadores de los Estados Unidos de Norteamérica y de cuatro ejecutivos de alto nivel de importantes corporaciones norteamericanas. Las cifras correspondientes a los gobernadores eran 44, 36, 52 y 40. Las cifras correspondientes a los ejecutivos eran 32, 60, 48 y 36. Calcule la media y el desvío estándar de los gobernadores y de los ejecutivos, y explique su cálculo a una persona que nunca ha asistido a un curso de estadística. Observe también de qué forma difieren las medias y los desvíos estándar e intente explicar el significado de esas diferencias, suponiendo que estos casos representan a los gobernadores y a los ejecutivos de grandes corporaciones de los Estados Unidos de Norteamérica en general.

4. Un estudio mide la cantidad de días que 216 empleados de una gran empresa faltaron a su trabajo durante el año anterior a la medición. Como parte de los resultados el investigador informa lo siguiente: "La cantidad de días de ausentismo durante el año anterior ($M = 9{,}21$; $SD = 7{,}34$) fue..." Explique la información contenida entre paréntesis a una persona que nunca ha asistido a un curso de estadística.

5. En una medición estándar de capacidad auditiva, la media es 300 y el desvío estándar es 20. a) Calcule las puntuaciones Z de las personas que presentaron registros de 340, 310 y 260. b) Calcule las puntuaciones brutas de aquellas personas cuyas puntuaciones Z, en esta prueba, fueron 2,4, 1,5, 0 y -4,5.

6. La puntuación de una persona en una prueba de aptitud verbal es de 81, y de 6,4 en una prueba de aptitud numérica. En el caso de la prueba de aptitud verbal, la media para las personas en general es 50 y el desvío estándar es 20. En el caso de la prueba de aptitud numérica, la media para las personas en general es 0 y el desvío estándar es 5. ¿Cuál es la mayor aptitud de esta persona, la verbal o la numérica? Explique su respuesta a una persona que nunca ha asistido a un curso de estadística.

APÉNDICE DEL CAPÍTULO: FÓRMULAS DE CÁLCULO OPTATIVAS DE LA VARIANZA Y EL DESVÍO ESTÁNDAR

Se han desarrollado formas alternativas pero matemáticamente equivalentes a las fórmulas de varianza y desvío estándar con el fin de facilitar los cálculos cuando se realizan a mano, tal como era necesario antes de la invención de la computadora o de la calculadora con función para el cálculo del desvío.

Como mencionamos anteriormente, ésta es la fórmula de definición para el cálculo de la varianza:

$$SD^2 = \frac{\Sigma(X-M)^2}{N} = \frac{SS}{N}$$

Es muy tedioso realizar este cálculo a mano, ya que primero deberíamos calcular el desvío de cada caso. Sin embargo, el numerador de esta ecuación, la suma de cuadrados, puede manipularse algebraicamente de modo tal que sólo sea necesario utilizar la suma de todas las observaciones (algo que ya habrá sido calculado para encontrar la media) y la suma de los cuadrados de cada observación real (lo que resulta mucho más rápido de calcular que tener que encontrar primero cada desvío y luego elevarlo al cuadrado). Esta fórmula alternativa es la siguiente:

$$SD^2 = \frac{\Sigma X^2 - (\Sigma X)^2/N}{N} \qquad (2\text{-}9)$$

Observemos que ΣX^2 implica que se eleva al cuadrado cada observación y luego se suman esos cuadrados. Por otro lado, $(\Sigma X)^2$ implica que primero se suman todas las observaciones y luego se eleva esa suma al cuadrado.

La fórmula de cálculo del desvío estándar es la raíz cuadrada de la fórmula de cálculo de la varianza:

$$SD = \sqrt{\frac{\Sigma X^2 - (\Sigma X)^2/N}{N}}$$

La tabla 2-8 muestra el cálculo de la varianza y el desvío estándar de la información correspondiente a nuestro ejemplo sobre sesiones de terapia, utilizando la fórmula de cálculo. Compare este cálculo con el que aparece en la tabla 2-2, que se basa en la misma información pero utiliza la fórmula de definición.

Tabla 2.8.
Cálculo de la varianza y el desvío estándar correspondiente al ejemplo sobre sesiones de terapia, utilizando las fórmulas de cálculo.

Cantidad de sesiones (X)	Cantidad de sesiones al cuadrado (X^2)
7	49
8	64
8	64
7	49
3	9
1	1
6	36
9	81
3	9
8	64
Σ: 60	426

$$SD^2 = \frac{\Sigma X^2 - (\Sigma X)^2/N}{N} = \frac{426 - 60^2/10}{10} = \frac{426 - 3.600/10}{10} = \frac{426 - 360}{10} = \frac{66}{10} = 6{,}6$$

$$SD = \sqrt{SD^2} = \sqrt{6{,}6} = 2{,}57$$

3 Correlación

Descripción del capítulo

- Variable independiente o predictora y variable dependiente.
- Cómo graficar correlaciones: diagrama de dispersión.
- Patrones de correlación.
- Cálculo de un índice del grado de correlación lineal: coeficiente de correlación de Pearson.
- Integración de los distintos pasos. Otros ejemplos.
- Prueba de la significación estadística del coeficiente de correlación.
- Cuestiones relacionadas con la interpretación del coeficiente de correlación.
- Controversias y desarrollos recientes: ¿qué es una gran correlación?
- Coeficientes de correlación según se describen en publicaciones científicas.
- Resumen.
- Términos clave.
- Ejercicios.
- Apéndice I del capítulo: fórmula de cálculo optativa para el coeficiente de correlación.
- Apéndice II del capítulo: prueba de hipótesis y su potencia sobre el coeficiente de correlación.

En un estudio realizado recientemente (Aron & Fraley, 1998), 84 alumnos universitarios, que en el momento del estudio mantenían relaciones amorosas, completaron varios cuestionarios. Uno de los cuestionarios era una escala de intimidad (basada en Sternber, 1986) que incluía ítems tales como "mantengo una relación de mutuo entendimiento con mi pareja" y "recibo considerable contención emocional de mi pareja". Otro cuestionario era una escala de idealización (basada en Murray, 1998) que preguntaba a los alumnos cómo calificaban a sus parejas en cuanto a "ingenio y ocurrencia", "inteligencia", "seguridad en sí mismo", y así sucesivamente. Investigaciones previas habían mostrado que la escala de idealización indica en qué medida un individuo percibe que su pareja posee esas características positivas **independientemente del grado en que su pareja realmente las posea** (según las medidas obtenidas a través del informe propio de la pareja en cuestión).

Uno de los resultados del estudio fue que cuánto mayor era la intimidad que un individuo creía tener con su pareja, más la idealizaba. Es decir, en general, los estudiantes que presentaban registros altos en la escala de intimidad también presentaban registros altos en la de idealización. Los estudiantes que presentaban registros bajos en la escala de intimidad tendían a presentar registros bajos en la escala de idealización.

También podemos observar este patrón visualmente. La figura 3-1 muestra el gráfico de los resultados arrojados por el estudio. Los valores de la escala de idealización se encuentran en el eje vertical; los valores de la escala de intimidad se encuentran en el eje horizontal. Ambos valores observados en cada estudiante se representa con un punto. El patrón general muestra que los puntos se ubican desde el ángulo inferior izquierdo hacia el ángulo superior derecho. Es decir, los valores bajos en una variable en general coinciden con valores bajos en la otra variable, y los valores altos con los altos. Si bien el patrón está lejos de reflejar una coincidencia uno a uno, puede observarse una clara tendencia general.

Este patrón de valores altos en una variable que coinciden con los valores altos en la otra variable, y bajos que coinciden con bajos, y moderados con moderados, es un ejemplo de **correlación**.

Existen innumerables ejemplos de correlación: en el caso de los niños, existe correlación entre la edad y la capacidad de coordinación; con respecto a los estudiantes, generalmente suponemos que existe correlación entre la cantidad de tiempo de estudio y la cantidad aprendida; en cuanto al mercado, usualmente suponemos que existe correlación entre precio y calidad, que los precios elevados coinciden con la buena calidad y los precios bajos con la mala calidad.

Este capítulo explora la naturaleza de la correlación, la forma de describirla gráficamente, los diferentes tipos de correlación, la forma de calcular el coeficiente de correlación (una medida del grado de correlación) y otros temas relacionados con la interpretación de un coeficiente de correlación. En el capítulo 4 tratamos el modo en que se utiliza la correlación para predecir el valor de una persona con respecto a una variable basándonos en el valor de esa persona en relación con otra variable (por ejemplo, predecir las calificaciones de una persona en la facultad sobre la base de sus calificaciones en la escuela secundaria). Al abordar los temas de la correlación y de la predicción, pasamos de la estadística que trata una sola variable (capítulos 1 y 2) a la estadística referida a la relación entre dos o más variable.

VARIABLES INDEPENDIENTES O PREDICTORAS Y VARIABLES DEPENDIENTES

Sin embargo, antes de dedicarnos al tema de la correlación es necesario presentar algunos términos importantes. Al estudiar la relación entre dos variables, frecuentemente pensamos en una variable como la **causa** y en la otra variable como el **efecto**. Por ejemplo, podríamos considerar a la intimidad como causa de la idealización. La variable considerada causa se denomina **variable independiente**, y la considerada efecto se denomina **variable dependiente**. (La variable dependiente se denomina así porque su valor depende del valor de la variable independiente. La variable independiente, por el contrario, se denomina de ese modo porque su valor no depende de la variable dependiente. En este pequeño mundo formado por sólo dos variables, en el que una es la causa de la otra, una es independiente y la otra es dependiente de la primera). En nuestro ejemplo, la intimidad (causa) sería la variable independiente y la idealización (efecto) la variable dependiente.

Sin embargo, tanto en el ejemplo mencionado como en muchos casos en psicología, es posible invertir la variable considerada causa y la variable considerada efecto. Fuera del laboratorio de pruebas, idealizar a nuestra pareja, por ejemplo, podría hacernos sentir que la relación es íntima, del mismo modo en que el hecho de sentir que la relación es íntima podría hacernos idealizar a nuestra pareja. La principal excepción aparece en experimentos reales, en los que el experimentador controla el nivel de la variable independiente; por ejemplo, asignando personas de manera aleatoria a diferentes niveles de esa variable (*véase* apéndice A).

Muchas veces a los investigadores no les agrada utilizar los términos **variable independiente y variable dependiente** en estudios en los que dos variables se miden, simplemente, tal como se manifiestan en un grupo de personas (como ocurre en nuestro ejemplo de intimidad e idealización). Sin embargo, aun cuando no podamos determinar con certeza cuál es la causa y cuál el efecto, es posible utilizar el conocimiento sobre una variable para **predecir** los valores correspondientes a la otra variable. Por ejemplo, según los hallazgos realizados en el estudio sobre intimidad e idealización, es razonable llegar a la conclusión de que es probable que las personas que sienten una relación especialmente íntima con sus parejas también las idealicen de una forma especial. En ese caso, estamos utilizando la intimidad para predecir la idealización. No importa realmente cuál es la causa o efecto subyacente, siempre que intimidad e idealización estén firmemente relacionadas.

Por lo tanto, al analizar dos variables relacionadas, algunos investigadores prefieren llamar a aquella a partir de la cual realizan la predicción, **variable predictora**. Sin embargo, la otra variable generalmente continúa denominándose variable dependiente. (El término adecuado para la variable predicha es **variable de criterio**, pero este término rara vez se utiliza en psicología, excepto en algunos textos sobre estadística). Según lo acostumbrado, en nuestro libro generalmente nos referiremos a una de las dos variables correlacionadas como la variable predictora, y a la otra como la variable dependiente.

Más adelante, en este mismo capítulo, profundizaremos el tema de la causalidad, y en el capítulo 4 veremos exclusivamente la predicción. Presentamos estos temas ahora en forma concisa, porque al estudiar la correlación es útil poder emplear los nombres de las dos variables que se correlacionan, y los nombres que hemos dado son los nombres convencionales. De hecho, es frecuente que en el campo de la psicología lo que nos interese sea sólo el grado en el que dos variables se relacionan. Probablemente no sea necesario dar a una variable ninguna condición especial como causa de la otra o como base para realizar predicciones sobre la otra. Sin embargo, a menudo los psicólogos se ven forzados a realizar tales distinciones aunque tengan que hacerlo en forma arbitraria, simplemente para cumplir los estándares establecidos para la creación de gráficos (como veremos muy pronto). Lo mismo sucede con la correlación múltiple, que es más compleja y que trataremos en el capítulo 4.

CÓMO GRAFICAR CORRELACIONES: DIAGRAMA DE DISPERSIÓN

La figura 3-1 representa la correlación entre intimidad e idealización: es un ejemplo de **diagrama de dispersión**. Un diagrama de dispersión permite observar a simple vista el grado y el patrón de relación entre las dos variables.

Cómo confeccionar un diagrama de dispersión

La creación de un diagrama de dispersión puede dividirse en tres pasos:

1. Dibujar los ejes y determinar qué variable se representa en cada uno de ellos. La variable independiente o predictora se ubica en el eje horizontal, la variable dependiente en el vertical. En la figura 3-1 ubicamos la variable intimidad en el eje horizontal y la variable idealización en el vertical. Lo hicimos de ese modo porque en el estudio estábamos interesados en observar si el grado de intimidad podría ser causa del grado de idealización.

2. Determinar la serie de valores que se van a utilizar para cada variable y marcarla en los ejes. Los números deben ir ascendiendo en cada eje, a partir del punto en el que los dos ejes se cortan. Comúnmente, comenzamos con el valor 0 ó con el menor valor que pueda tener la medida, y ascendemos gradualmente hasta llegar al mayor valor posible de esa medida. Cuando no exista un valor mínimo o máximo posible, que sea evidente o razonable, comenzamos o terminamos la serie con el valor mínimo o máximo presentado comúnmente por las personas que conforman el grupo de interés para el estudio. (Por otro lado, a diferencia del tipo de gráficos descriptos en el capítulo 1, un diagrama de dispersión no se dibuja siguiendo la relación de 1,5 a 1 para el an-

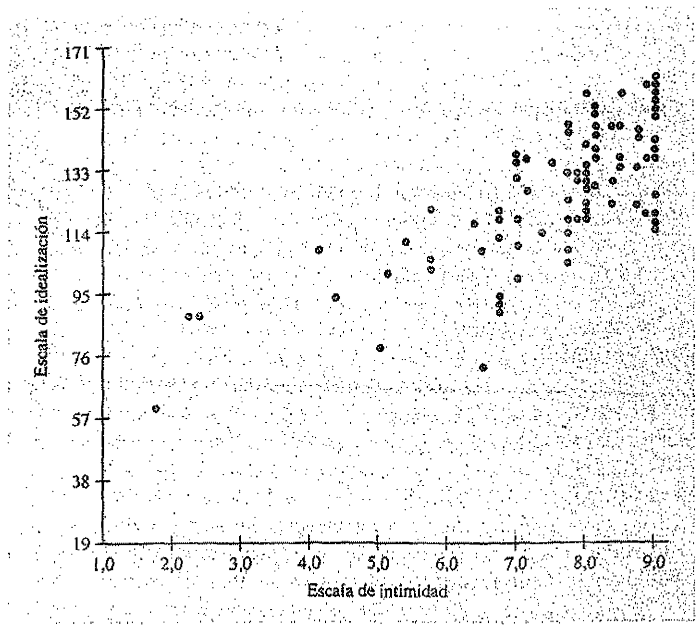


Figura 3-1. Diagrama de dispersión que muestra la correlación entre intimidad e idealización de 85 alumnos universitarios al describir sus actuales relaciones amorosas. (**Fuente**: Aron & Fraley, 1998).

cho y para la altura respectivamente. Los diagramas de dispersión son cuadrados, con una relación 1 a 1 para los ejes horizontales y verticales).

En la figura 3-1, el eje horizontal comienza con el valor 1 (el menor puntaje posible en la escala de intimidad, que es un promedio de varias preguntas contestadas cada una de ellas con referencia a una escala del 1 al 9). El eje vertical comienza con 19, que es la menor puntuación posible de la escala de idealización (esta escala incluye 19 ítems, clasificados del 1 al 9. El puntaje total de la escala es la suma de los 19 ítems). El valor más alto en el eje horizontal es 9,0, la máxima puntuación posible en la escala de intimidad. El valor más alto en el eje vertical es 171, la mayor puntuación posible en la escala de idealización.

3. Marcar un punto por el par de observaciones de cada persona. Ubicar el lugar en el eje horizontal que corresponde al valor observado de la persona en la variable predictora. Luego moverse hacia arriba hasta llegar a la altura en el eje vertical que corresponde al valor observado de la misma persona con respecto a esa variable, y marcar un punto bien claro.

Si en un mismo lugar coinciden dos casos, se puede escribir el número 2 en ese lugar o marcar un segundo punto lo más cerca posible del primero, si es posible tocándolo, pero dejando en claro que en realidad hay 2 puntos en el mismo lugar.

Ejemplo

Supongamos que una empresa está pensando aumentar la cantidad de personal bajo el mando de cada uno de sus gerentes de piso. Sin embargo, la empresa está preocupada por el estrés que ésto podría provocar a sus gerentes. La *empresa supone que cuantas más personas supervise* un gerente, mayor será el estrés sufrido por él. Para analizar la situación, un psicólogo laboral sugiere estudiar a cinco gerentes seleccionados al azar de entre todos los gerentes de piso de la empresa. (En la práctica, debería utilizarse un grupo mucho mayor, pero aquí utilizaremos sólo cinco casos para simplificar el ejemplo). Se entrega a cada uno de los cinco gerentes un cuestionario de medición de estrés en el cual los posibles registros van de 0 (estrés nulo) a 10 (estrés extremo). Los resultados podrían ser como los que indica la tabla 3-1.

1. Dibujar los ejes y determinar qué variable representa cada uno de ellos. La empresa está interesada en el efecto causado en el nivel de estrés por la cantidad de empleados supervisados. Por lo tanto, consideramos la cantidad de empleados supervisados como la variable predictora y ubicamos esa información en el eje horizontal; el nivel de estrés es la variable dependiente y, por lo tanto, debe ubicarse en el eje vertical. (*Véase* figura 3-2a).

2. Determinar la serie de valores que se van a utilizar para cada variable y marcarla en los ejes. Para el eje horizontal, supongamos que en esta empresa no se permite a ningún gerente supervisar más de 12 empleados. Por lo tanto, el eje horizontal va de 0 a 12. El eje vertical va de 0 a 10, que son los límites del cuestionario de medición de estrés. (*Véase* figura 3-2b).

3. Marcar un punto por el par de observaciones de cada persona. En el caso del primer gerente, la cantidad de empleados supervisados es 6. Localizamos el número 6 en el eje horizontal. Luego, subimos hasta alcanzar el nivel del número 7 en el eje vertical (el nivel de estrés del primer gerente). Marcamos un punto en ese lugar (*véase* figura 3-2c). Seguimos el mismo procedimiento con cada uno de los cuatro gerentes restantes. El resultado debería ser el que muestra la figura 3-2d.

PATRONES DE CORRELACIÓN

Hasta ahora hemos considerado aquellas situaciones en las que los valores altos coinciden con los altos, los bajos con los bajos y los medianos con los medianos. A ese tipo de situación se la denomina **correlación positiva**. Debido a que el patrón que muestra el diagrama de dispersión se aproxima a una línea recta, es también un ejemplo de **correlación lineal**.

Por ejemplo, en el diagrama de dispersión de la figura 3-1 se podría dibujar una recta que muestre la tendencia general de los puntos, tal como lo hemos hecho en la figura 3-3. Del mismo modo, se podría dibujar una recta en nuestro segundo ejemplo, como lo muestra la figura 3-4. (Una de las razones por las que estos casos de correlaciones lineales se denominan "positivas" se debe a que, en geometría, la pendiente de una recta es positiva cuando observamos que la recta se eleva a medida que desplazamos nuestra mirada desde la izquierda hacia la derecha del gráfico. En el capítulo 4, aprenderemos reglas precisas para trazar tales rectas y determinar su pendiente).

Tabla 3-1.
Empleados supervisados y nivel de estrés (datos ficticios).

Empleados supervisados	*Nivel de estrés según cuestionario*
6	7
8	8
3	1
10	8
8	6

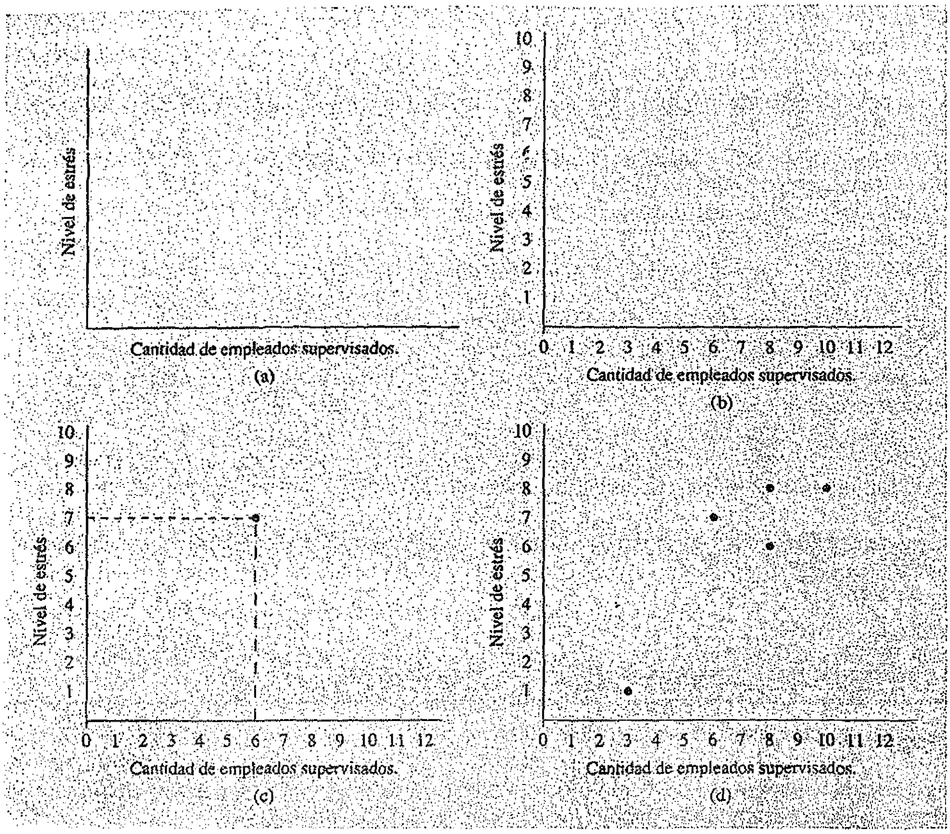


Figura 3-2. Cómo hacer un diagrama de dispersión. (a) Se determinan los ejes, la variable predictora (empleados supervisados) se ubica en el eje horizontal y la variable dependiente (nivel de estrés) en el eje vertical. (b) Se marca la serie valores sobre los ejes. (c) Se marca el punto determinado por el par de valores observados correspondientes al primer gerente. (d) Se marca un punto donde se cruza por cada par de valores observados de los cinco gerentes.

Correlaciones negativas

A veces, la relación entre las variables no es positiva. Por el contrario, los valores altos coinciden con los bajos y los bajos con los altos. A esto se denomina correlación negativa. Por ejemplo, en un estudio de las relaciones amorosas entre estudiantes (Aron & Fraley, 1998), los investigadores descubrieron que cuanto más lejos de su pareja vive una persona (en función de los minutos de viaje), menos cantidad de actividades comparte con su pareja. El diagrama de dispersión de la figura 3-5 representa gráficamente este patrón de correlación.

Incluimos una recta en la figura para remarcar la tendencia general de los puntos; así, podemos observar que a medida que la recta avanza hacia la derecha, también se dirige hacia abajo. Es decir, cuantas más son las horas de viaje, menos actividades se realizan en pareja.

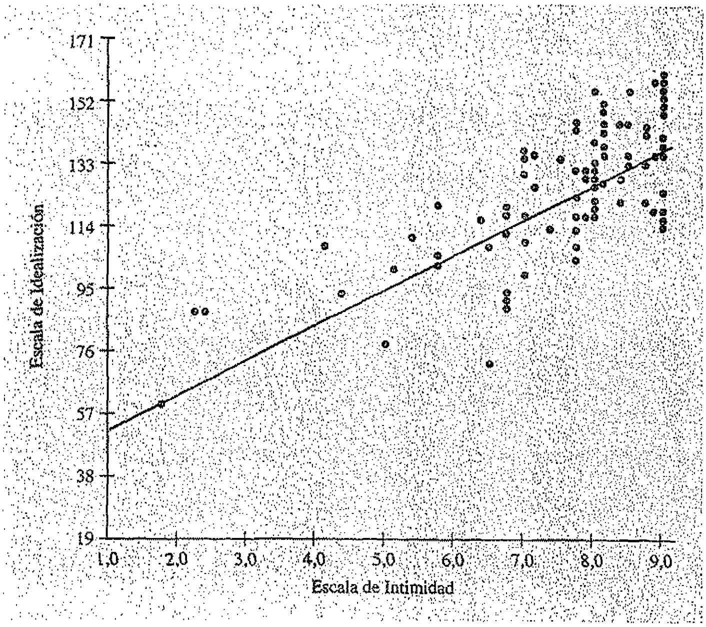


Figura 3-3. Diagrama de dispersión de la figura 3-1 con una recta que indica la tendencia general. (**Fuente:** Aron & Fraley, 1998).

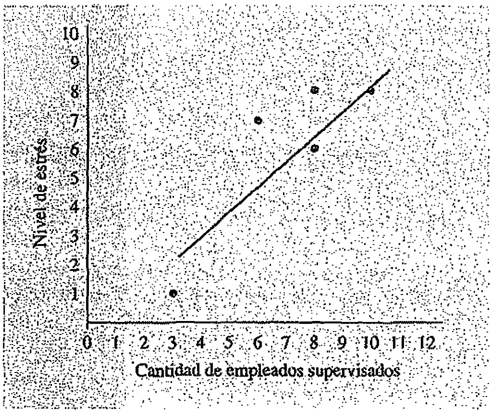


Figura 3-4. Diagrama de dispersión de la figura 3-2d con una recta que indica la tendencia general.

Una investigación realizada por Bardsley y Rhodes (1996), dos psicólogos especializados en organizaciones empresariales, ilustra también una correlación negativa. A través de un estudio realizado con 174 obreros, descubrieron que el hecho de llegar tarde a trabajar tenía una correlación lineal negativa con la satisfacción laboral. Cuanto mayor era el grado de satisfacción la-

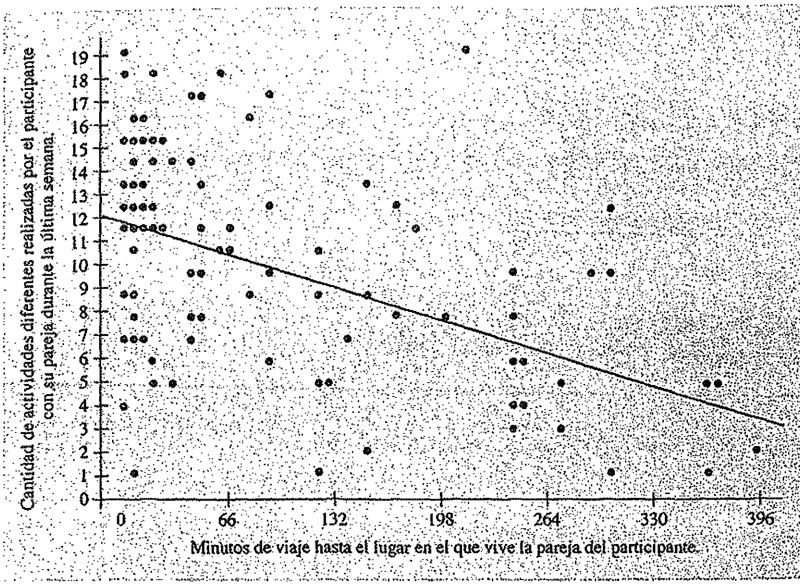


Figura 3-5. Diagrama de dispersión con una recta que indica la tendencia general de una correlación negativa entre dos variables: distancia en minutos de viaje y cantidad de actividades diferentes que el participante realiza con su pareja. (**Fuente**: Aron & Fraley, 1998).

boral de los obreros, menos frecuentemente llegaban tarde. En otras palabras, cuanto menor era el nivel de satisfacción de los obreros, con más frecuencia llegaban tarde a trabajar.

Correlaciones curvilíneas

En algunos casos, la relación entre dos variables no sigue una línea recta positiva o negativa, sino un patrón más complejo denominado **correlación curvilínea**. Por ejemplo, se sabe que hasta determinado nivel, una mayor ansiedad fisiológica hace que uno se desempeñe mejor en cualquier tarea (como por ejemplo, una prueba de matemática). A partir de ese nivel, una mayor ansiedad fisiológica hace que el rendimiento empeore. Es decir, desde estar casi dormido hasta un nivel moderado de ansiedad, la efectividad aumenta. Al superar ese nivel moderado, el aumento de la ansiedad puede "acelerar" demasiado a un individuo, impidiéndole tener un buen rendimiento. Este patrón curvilíneo en particular está representado en la figura 3-6, en donde se observa que sería imposible dibujar una línea recta para describirlo. La figura 3-7 muestra algunos otros ejemplos de relaciones curvilíneas.

A través del método usual de cálculo de la correlación (método que aprenderemos en este capítulo) obtenemos el grado de correlación lineal. Si el verdadero patrón de asociación es curvilíneo, calcular la correlación con el método usual podría dar como resultado muy poca correlación o una correlación nula. Por eso es muy importante observar los diagramas de dispersión para descubrir estas relaciones más interesantes, antes de realizar correlaciones automáticamente con el método usual, suponiendo que la única relación posible sea una línea recta.

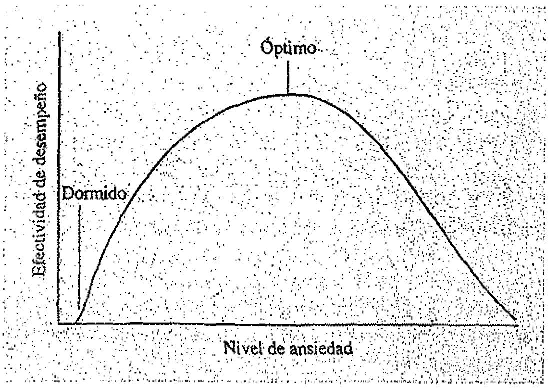


Figura 3-6. Ejemplo de relación curvilínea: desempeño en una tarea y ansiedad.

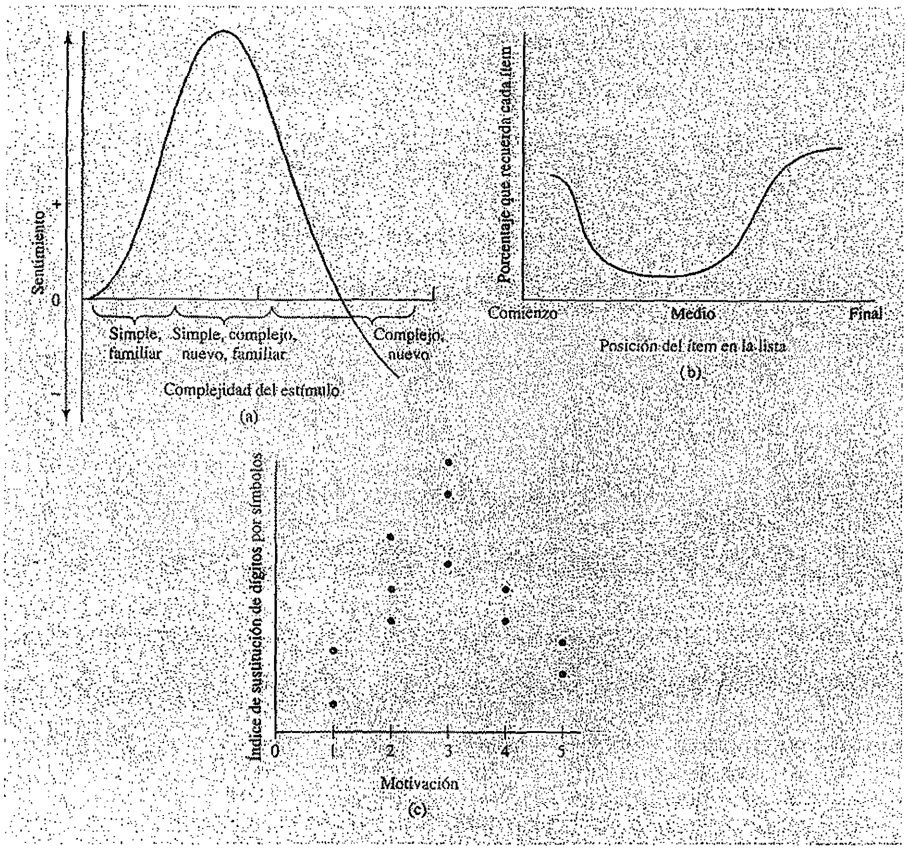


Figura 3-7. Ejemplo de relaciones curvilíneas: (a) modo en que uno se siente y complejidad del estímulo; (b) cantidad de personas que recuerdan un ítem y posición del mismo en una lista, y (c) índice de sustitución de dígitos por símbolos y motivación en niños.

Correlación nula

También es posible que no exista ningún tipo de relación entre dos variables. Por ejemplo, si hiciéramos un análisis entre la creatividad y el número de calzado, los resultados podrían ser similares a los que muestra la figura 3-8. Los puntos se dispersan en todas las direcciones, y no existe línea recta ni de ningún otro tipo que pueda considerarse indicio razonable de una tendencia. Esto simplemente es una **correlación nula**.

En investigaciones reales, algunas veces existe una relación entre dos variables, pero al no ser muy fuerte, es difícil notarla en un diagrama de dispersión. Esto suele suceder especialmente en análisis en los que se estudia a una gran cantidad de personas y la relación entre las dos variables es muy leve o subliminal. En esos casos, uno podría estar seguro de que la relación es más que una coincidencia, justamente debido a su leve pero consistente presencia en un grupo tan grande de personas. La figura 3-9 muestra un diagrama de dispersión con una leve correlación lineal positiva entre dos variables. ¿Es posible observarla?

Figura 3-8. Dos variables sin asociación entre sí, creatividad y número de calzado (datos ficticios).

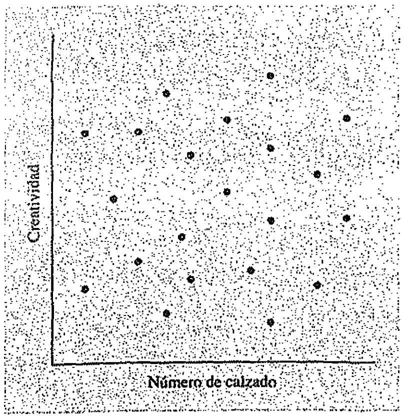
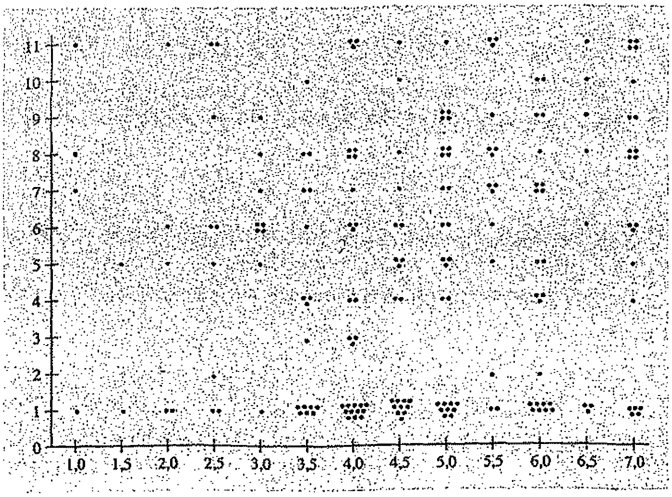


Figura 3-9. Diagrama de dispersión en el que se encontró una leve correlación lineal positiva entre las dos variables.



CÁLCULO DE UN ÍNDICE DEL GRADO DE CORRELACIÓN LINEAL: COEFICIENTE DE CORRELACIÓN DE PEARSON

Al observar un diagrama de dispersión obtenemos un indicio aproximado del tipo y grado de relación entre dos variables. Sin embargo, observar el gráfico no es un método muy preciso. Es necesario obtener un número que represente el grado exacto de correlación.

Grado de correlación

El **grado de correlación** indica en qué medida existe un patrón claro de alguna relación en particular entre dos variables. Por ejemplo, vimos que existe una correlación lineal positiva cuando los valores altos coinciden con los valores altos, los medios con los medios y los bajos con los bajos. Por lo tanto, el grado de una correlación de este tipo determina cuántos valores altos coinciden con otros también altos, y así sucesivamente. Del mismo modo, el grado de correlación lineal negativa indica cuántos valores altos de una variable coinciden con valores bajos de la otra, y así sucesivamente. En cuanto a los diagramas de dispersión, un alto grado de correlación lineal significa que todos los puntos se encuentran muy cerca de una línea recta (la recta que se inclina hacia arriba o hacia abajo según la correlación lineal sea positiva o negativa). Una correlación lineal perfecta es aquella en la que todos los puntos están ubicados exactamente sobre la línea recta.

Determinación del grado de correlación lineal

Lo primero que necesitamos para determinar el grado de correlación lineal es alguna forma de medir qué es un valor alto o bajo, y en qué medida es alto o bajo un determinado valor alto o bajo. (Otro método, descripto en el capítulo 4, consiste en determinar la distancia entre cada punto y la línea). El método aquí propuesto implica comparar valores de diferentes variables en forma coherente. Como vimos en el capítulo 2, la mejor forma de resolver el problema de comparar manzanas con naranjas es a través de las puntuaciones Z.

A modo de revisión, una puntuación Z es la cantidad de desvíos estándar a los que se encuentra una observación de la media. No importa la escala con la cual se haya realizado la medición; si uno convierte las puntuaciones originales en puntuaciones Z, el efecto es el mismo. Una puntuación original alta (es decir, que se encuentra por encima de la media de los otros valores de la variable) siempre tendrá una puntuación Z positiva, y una puntuación original baja (por debajo de la media) siempre tendrá una puntuación Z negativa. Más aún, cualquiera sea la medida usada en particular, las puntuaciones Z brindan un indicio estándar de cuán alta o baja es cada puntuación. Por ejemplo, una puntuación Z de 1 está siempre exactamente 1 desvío estándar por sobre la media, una puntuación Z de 2 está siempre dos desvíos estándar por sobre la media. Las puntuaciones Z de una variable son directamente comparables con las puntuaciones Z de otra variable.

Existe otra razón por la cual se utilizan puntuaciones Z para calcular el grado de correlación, y está relacionada con lo que sucede si se multiplica un valor de una variable por un valor de la otra variable, cálculo que se denomina **producto cruzado**. Cuando se utilizan puntuaciones Z, el cálculo se denomina **producto cruzado de puntuaciones Z**. Si se multiplica una puntuación Z alta por una puntuación Z alta, siempre se obtiene un producto cruzado positivo debido a que, más allá de qué escala se utilice, los valores que se encuentran por sobre la media (valores altos) se transforman en puntuaciones Z positivas, y un número positivo multiplicado por un número positivo siempre produce un número positivo. Más aún, y esto es lo interesante, si se multiplica una puntuación Z baja por otra puntuación Z baja, también se obtiene siempre un producto cruzado positivo debido a que, más allá de qué escala se utilice, los valores que se encuentran por debajo de la media (valores bajos) se convierten en puntuaciones Z negativas, y un número negativo multiplicado por un número negativo siempre produce un número positivo.

Cuando los valores altos de una variable coinciden con los altos de la otra, y los bajos de una con los bajos de la otra, el producto cruzado de puntuaciones Z siempre será positivo. Supongamos que tomamos toda una distribución de valores observados y multiplicamos la puntuación Z de cada persona en una variable por la puntuación Z de esa persona en la otra variable. El resultado de esta operación, cuando los valores altos coinciden con los altos y los bajos con los bajos, es que todos los productos resultarán positivos. Si luego sumamos esos productos cruzados de puntuaciones Z de todas las personas incluidas en el estudio, que son todos positivos, obtendremos un gran número positivo.

Por otro lado, consideremos lo que sucedería con una correlación lineal negativa, en la que los valores altos coinciden con los bajos y los bajos con los altos. En cuanto a las puntuaciones Z, esto significaría multiplicar positivos con negativos y negativos con positivos, y obtendríamos todos productos cruzados negativos que, sumados, darían como resultado un gran número negativo.

Finalmente, supongamos que no existe una correlación lineal. En esa situación, en algunos casos los valores altos de una variable coincidirían con valores altos de la otra variable (y algunos valores bajos coincidirían con valores bajos), dando como resultado productos cruzados positivos. En otros casos, los valores altos de una variable coincidirían con valores bajos de la otra (y algunos valores bajos coincidirían con valores altos), dando productos cruzados negativos. Al sumar los productos cruzados de todas las personas incluidas en el estudio, los productos cruzados positivos y negativos se cancelarían unos a otros, dando un total cercano a 0.

En cada una de las situaciones arriba mencionadas convertimos todos los valores en puntuaciones Z, multiplicamos las dos puntuaciones Z de cada persona y sumamos los productos cruzados. El resultado es un gran número positivo, si existe una correlación lineal positiva; un gran número negativo, si existe una correlación lineal negativa, y un número cercano a 0, si no existe correlación lineal.

Sin embargo, todavía no hemos resuelto el problema de determinar el grado de una correlación positiva o negativa. Aparentemente, cuanto mayor sea el número, mayor será la correlación. ¿Pero a partir de qué nivel se considera que un número grande es tal, y qué números grandes no lo son tanto? No podemos definirlo simplemente por la suma de los productos cruzados, que aumenta sólo por el hecho de incluir más participantes en el estudio. (Es decir, un estudio con 100 participantes presentaría una suma mayor de productos cruzados que el mismo estudio con sólo 25 participantes).

La solución a este problema es dividir la suma de productos cruzados de puntuaciones Z por la cantidad de casos. Es decir, calculamos el **promedio de los productos cruzados de puntuaciones Z**, que nunca podrá ser mayor a +1, y que en el caso de ser igual a +1, indicaría una **correlación perfecta** lineal positiva. Por otro lado, el valor mínimo de este promedio es −1, e indicaría una correlación perfecta lineal negativa. En el caso de que no exista correlación lineal, el promedio de los productos cruzados de puntuaciones Z será 0.

Generalmente, las correlaciones no son perfectas. En el caso de una correlación lineal positiva que no es perfecta, el promedio de los productos cruzados de puntuaciones Z estará entre 0 y +1. Para decirlo de otro modo, si la tendencia general de las puntuaciones es ascendente y hacia la derecha, pero no coinciden exactamente con la línea recta, este número estará entre 0 y +1. La misma regla se aplica para las correlaciones negativas: el número estará entre 0 y −1.

Cuadro3-1.
Galton: un caballero genial.

Francis Galton es considerado el inventor del cálculo estadístico denominado correlación, aunque Karl Pearson y otros desarrollaron las fórmulas. En el capítulo 14 (cuadro 14-1) aprenderemos algo sobre Pearson, y en el capítulo 9 (cuadro 9-1) sobre William S. Gosset, otro importante estadístico precursor, inventor de la prueba *t*. Gosset era alumno y colega de Karl Pearson. Pearson, a su vez, fue alumno y colega de Galton (a quien Pearson otorgó todo el crédito por el descubrimiento de la correlación). En otras palabras, el mundo de la estadística en ese momento pertenecía a un pequeño y selecto club británico (*véase* cuadro 16-1). De hecho, casi toda la ciencia en general era una especie de club apenas más amplio. Por ejemplo, Galton estaba muy influenciado también por su propio primo, Charles Darwin.

De todos los miembros de este club, Galton era tal vez el más típico de su época, un caballero científico, excéntrico y acaudalado. Además de su trabajo en estadística, poseía un título en medicina, exploró el "África oscura", inventó anteojos para leer debajo del agua, experimentó con mapas estereoscópicos, incursionó en la meteorología y antropología y escribió un informe sobre la recepción de señales inteligibles desde las estrellas.

Sobre todo, Galton era un contador compulsivo. Algunos de sus recuentos eran infames. Una vez, mientras asistía a una cátedra, contó las veces por minuto en que la audiencia se impacientaba, buscando variaciones relacionadas con lo aburrido del tema tratado. En dos oportunidades se hizo retratar y se dedicó a contar las pinceladas que daba el artista por hora, llegando a la conclusión de que cada retrato requería un promedio de 20.000 pinceladas. Y mientras caminaba por las calles de varias ciudades de las Islas Británicas, clasificó la belleza de los habitantes femeninos pulsando un mecanismo de grabación que llevaba en su bolsillo con el que registraba "bueno", "regular" o "malo".

Sin embargo, el interés que consumía a Galton era contar la cantidad de genios, criminales y otros caracteres típicos existentes en las distintas familias. Quería comprender cómo se producía cada tipo para que la ciencia pudiera mejorar la raza humana: la finalidad sería estimular a los gobiernos a imponer la eugenesia, es decir, la procreación selectiva orientada a lograr un mayor nivel de inteligencia, comportamiento moral adecuado y otras cualidades, que serían determinadas, por supuesto, por el eugenista. (A partir de esa época, la eugenesia cayó en el descrédito). El concepto de correlación surgió directamente de los primeros y sencillos esfuerzos de Galton en ese sentido: el estudio de la relación entre la altura de los niños y sus padres.

En realidad, gran parte de la ciencia estadística, o "biometría", como la denominaba Galton, surgió como aplicación de la matemática a temas relacionados con la biología y las ciencias sociales. Y de todos los cálculos estadísticos, la correlación fue una herramienta especialmente útil para estas ciencias en las que, por lo general, no podían realizarse experimentos rigurosos tales como los experimentos de procreación en humanos. Al principio, el método de Galton para medir la tendencia con que "una cosa ocurría junto con otra" parecía ser casi el mismo que para probar la causa de algo. Por ejemplo, si podía demostrarse matemáticamente que la mayoría de las personas más brillantes provenían de unas pocas familias británicas de alta alcurnia, y la mayoría de las personas menos inteligentes provenían de las familias pobres, aparentemente se probaría que la inteligencia era producto de la herencia de ciertos genes (siempre que uno fuera lo suficientemente prejuicioso como para pasar por alto las diferencias en cuanto a oportunidades educativas). El mismo estudio podría probar en forma más convincente que si uno era miembro de una de las mejores familias británicas, la historia lo convertiría en un excelente ejemplo de la facilidad con que se malinterpreta el significado de la correlación.

Referencias: Peters (1987); Tankard (1984).

El coeficiente de correlación

El promedio de los productos cruzados de puntuaciones Z es, entonces, un excelente modo de calcular el grado de correlación lineal. Se lo denomina **coeficiente de correlación.** También se lo llama **coeficiente de correlación de Pearson** (o, para ser muy tradicionales, **coeficiente de correlación producto-momento de Pearson**). Lleva el nombre de Karl Pearson (a quien presentaremos en el cuadro 14-1). Pearson, junto con Francis Galton (*véase* cuadro 3-1), desempeñó un papel fundamental en el desarrollo del coeficiente de correlación. El coeficiente de correlación se representa con la letra *r*, que es la forma abreviada de **regresión**, un concepto muy relacionado con la correlación (que veremos en el capítulo 4). También es importante saber que en algunas publicaciones científicas se hace referencia a los coeficientes de correlación como **correlaciones de orden cero** (veremos las razones para este nombre en el capítulo 17).

La figura 3-10 muestra diagramas de dispersión e indica el coeficiente de correlación de varios ejemplos.

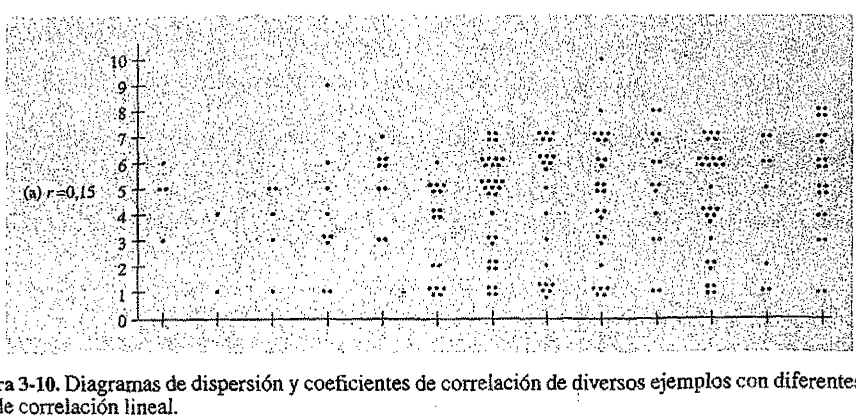


Figura 3-10. Diagramas de dispersión y coeficientes de correlación de diversos ejemplos con diferentes grados de correlación lineal.

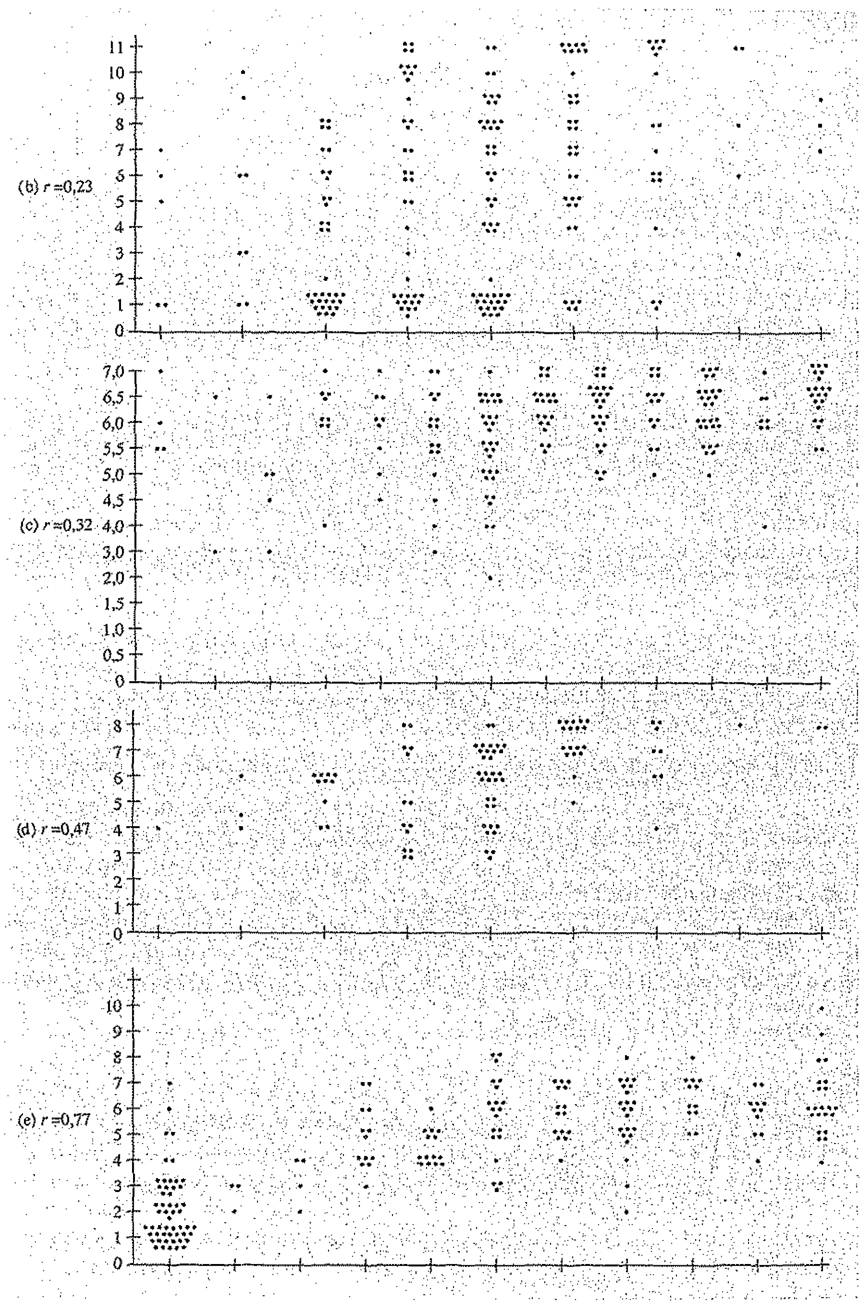


Figura 3-10. (continuación)

Fórmula del coeficiente de correlación

La exposición precedente puede resumirse en unos pocos símbolos, la fórmula del coeficiente de correlación:

$$r = \frac{\Sigma Z_X Z_Y}{N} \tag{3-1}$$

r es el coeficiente de correlación, Z_X es la puntuación Z de cada persona en la variable X, Z_Y es la puntuación Z de cada persona en la variable Y. $Z_X Z_Y$ es igual al producto Z_X por Z_Y (el producto cruzado de puntuaciones Z) de cada persona, y $\Sigma Z_X Z_Y$ es la suma de los productos cruzados de todas las personas incluidas en el estudio. N es la cantidad de personas que participan en el estudio. Uniendo todos los datos, $\Sigma Z_X Z_Y$ dividida por N, es el promedio de los productos cruzados de puntuaciones Z.

Pasos que se deben seguir para calcular el coeficiente de correlación

Los cuatro pasos que se deben seguir para calcular el coeficiente de correlación son los siguientes:

1. Convertir todas las observaciones en puntuaciones Z. Para esto es necesario calcular la media y el desvío estándar de cada variable y luego la puntuación Z correspondiente a cada valor observado.
2. Calcular el producto cruzado de las puntuaciones Z de cada persona. Es decir, por cada persona, multiplicar la puntuación Z en una variable por la puntuación Z en la otra variable.
3. Sumar todos los productos cruzados de puntuaciones Z.
4. Dividir el resultado por la cantidad de personas que participan en el estudio.

Fórmula de definición versus fórmula de cálculo para el coeficiente de correlación

El procedimiento que acabamos de describir, basado en la fórmula de definición, esclarece la lógica implícita en el cálculo del coeficiente de correlación. Seguir estos pasos en los ejercicios que presentamos a continuación (y en los ejercicios al final del capítulo) ayuda a incorporar la lógica mencionada. Sin embargo, para calcular el coeficiente de correlación en un estudio real, casi siempre usaríamos una computadora. La fórmula de cálculo que aparece en el apéndice I de este capítulo facilitará mucho el trabajo, si alguna vez fuera realmente necesario. En un estudio real con muchos registros, calcular un coeficiente de correlación manualmente (o con una calculadora).

Ejemplo

Intentemos aplicar los pasos enumerados al ejemplo del nivel de estrés de los gerentes.

1. Convertir todas las observaciones en puntuaciones Z. Comenzando con la cantidad de empleados supervisados, la media es 7 (la suma, que es igual a 35, dividida por 5 gerentes) y el desvío estándar es 2,37 (la suma de los desvíos cuadráticos, 28, dividida por 5 gerentes, es igual a una varianza de 5,6, cuya raíz cuadrada es 2,37). En el caso del primer gerente, entonces, un valor observado de 6 es una unidad por debajo de la media 7, y 1 dividido 2,37 es 0,42. Por lo tanto, la puntuación Z del primer gerente referido a la cantidad de empleados supervisados se ubica a 0,42 desvíos estándares por debajo de la media o, lo que es igual, presenta una puntuación Z de – 0,42.

Calculamos el resto de las puntuaciones Z del mismo modo y las ordenamos en las columnas correspondientes de la tabla 3-2.

2. Calcular el producto cruzado de las puntuaciones Z de cada persona. En el caso del primer gerente, multiplicamos – 0,42 por 0,38: el resultado es –0,16. La última columna de la tabla 3-2 muestra los productos cruzados de todos los gerentes.

3. Sumar los productos cruzados de puntuaciones Z. Como lo indica la tabla 3-2, el total es 4,38.

4. Dividir el resultado del paso anterior por la cantidad de personas incluidas en el estudio, es decir, 4,38 dividido 5 (la cantidad de gerentes incluidos en el estudio). El resultado es 0,876. Este es el coeficiente de correlación que, redondeado, es igual a 0,88. Aplicando la fórmula del coeficiente de correlación,

$$r = \frac{\Sigma Z_X Z_Y}{N} = \frac{4{,}38}{5} = 0{,}88$$

Dado que el coeficiente de correlación calculado es positivo y cercano a 1, es decir, el mayor valor posible, podemos afirmar que estamos frente a una correlación lineal fuertemente positiva.

INTEGRACIÓN DE LOS DISTINTOS PASOS. OTROS EJEMPLOS

En general, cuando nos encontramos frente a un problema relacionado con la correlación, el método adecuado consiste en hacer primero un diagrama de dispersión. Luego, si el diagrama de dispersión no muestra un patrón curvilíneo claro, se procede a calcular el coeficiente de correlación. Aun cuando no exista un patrón curvilíneo, es conveniente observar un poco más en detalle el diagrama de dispersión. La idea es estimar en forma aproximada el grado y la dirección de la correlación lineal, como forma de control en el caso de que haya errores al calcular concretamente el coeficiente de correlación.

Tabla 3-2 .
Cálculo del coeficiente de correlación para el ejemplo del nivel de estrés de los gerentes (datos ficticios).

Cantidad de Empleados Supervisados (X)				Nivel de Estrés (Y)			Prod. Cruzados	
X	$X-M$	$(X-M)^2$	Z_X	Y	$Y-M$	$(Y-M)^2$	Z_Y	$Z_X Z_Y$
6	–1	1	–0,42	7	1	1	0,38	–0,16
8	1	1	0,42	8	2	4	0,77	0,32
3	–4	16	–1,69	1	–5	25	–1,92	3,24
10	3	9	1,27	8	2	4	0,77	0,98
8	1	1	0,42	6	0	0	0,00	0,00
$\Sigma = 35$		$SS = 28$		$\Sigma = 30$		$SS = 34$		$\Sigma Z_X Z_Y = 4{,}38$
$M = 7$		$SD^2 = 5{,}60$		$M = 6$		$SD^2 = 6{,}80$		$r = 0{,}88$
		$SD = 2{,}37$				$SD = 2{,}61$		

Resumen de los pasos a seguir en un problema de correlación

Combinando los distintos procedimientos tratados en este capítulo, los pasos a seguir son los siguientes:

1. Construir un diagrama de dispersión.
   a) Dibujar los ejes y determinar qué variable va en cada uno de ellos.
   b) Determinar la serie de valores que se van a utilizar para cada variable y marcarla en los ejes.
   c) Marcar un punto por el par de observaciones de cada persona.
2. Determinar si el patrón es claramente curvilíneo. Si lo es, no se calcula el coeficiente de correlación (o si se lo calcula, debe tenerse en cuenta que sólo se está describiendo el grado de relación lineal).
3. Estimar la dirección y el grado de correlación lineal.
4. Calcular el coeficiente de correlación.
   a) Convertir todas las observaciones en puntuaciones Z.
   c) Calcular el producto cruzado de las puntuaciones Z de cada persona.
   d) Sumar los productos cruzados de puntuaciones Z.
   e) Dividir el resultado por la cantidad de personas incluidas en el estudio.
5. Controlar el signo y el tamaño del coeficiente de correlación calculado, comparándolo con la estimación visual realizada a partir del diagrama de dispersión.

Como observamos anteriormente, en una investigación real el coeficiente de correlación se calcularía en cuestión de segundos utilizando una computadora. La mayoría de los paquetes de computación pueden realizar también un diagrama de dispersión (y algunos hasta explican cómo se establecen sus ejes y escalas). Sin embargo, la finalidad de incluir los ejemplos (y resolver los ejercicios al final del capítulo) y aplicar todos estos pasos bastante tediosos es comprender la lógica implícita en los resultados que, con tanta facilidad, surgen de la computadora.

Ejemplo

Supongamos que una persona que investiga el funcionamiento de la memoria realiza un experimento para comprobar la teoría de que la cantidad de exposiciones a una palabra aumenta las probabilidades de que sea recordada. Dos individuos son elegidos al azar para observar una lista de 10 palabras una sola vez, otros dos individuos observan la lista dos veces, y así sucesivamente, hasta llegar a ocho exposiciones de cada palabra, y 16 participantes en total. La tabla 3-3 indica los resultados de este experimento ficticio. (Un estudio real de este tipo probablemente daría un resultado más curvilíneo debido a que, en esta clase de investigaciones, cuanto mayor sea la cantidad de exposiciones, menor será el aumento relativo de palabras recordadas).

1. Construir un diagrama de dispersión.
   a) Trazar los ejes y determinar qué variable deberá marcarse en cada uno de ellos. Según el diseño del experimento, la cantidad de exposiciones es la variable independiente, por lo que estará ubicada en el eje horizontal. La cantidad de palabras recordadas es la variable dependiente, por lo que estará ubicada en el eje vertical (*véase* figura 3-11a).
   b) Determinar la serie de valores que se van a utilizar para cada variable, y luego marcarla en los ejes. En el estudio que estamos analizando, la cantidad de exposiciones varía de 1 a 8, pero comenzaremos con 0 para cumplir con las reglas convencionales. La cantidad de palabras recordadas no puede ser menor que 0 ni mayor que 10, cantidad total de palabras en la lista (*véase* figura 3-11b).

Tabla 3-3.
Efecto del número de exposiciones en la cantidad de palabras recordadas.

Número de identificación	Cantidad de exposiciones	Cantidad de palabras recordadas
1	1	4
2	1	3
3	2	3
4	2	5
5	3	6
6	3	4
7	4	4
8	4	6
9	5	5
10	5	7
11	6	2
12	6	9
13	7	6
14	7	8
15	8	9
16	8	8

c) Marcar los puntos determinados por el par de observaciones de cada persona. El primer punto se ubica con coordenada 1 según el eje horizontal, y 4 según el eje vertical. Marcando cada uno de los puntos, de este mismo modo, completamos el diagrama de dispersión (*véase* figura 3-11c).

2. Determinar si el diagrama es claramente curvilíneo. Parece existir una fuerte tendencia lineal.
3. Estimar la dirección y el grado de correlación lineal. Los puntos van hacia arriba y hacia la derecha, y la mayoría de ellos están ubicados muy cerca de una línea recta imaginaria. Por lo tanto, aparentemente se trata de una correlación lineal positiva bastante fuerte.
4. Calcular el coeficiente de correlación.

a) Convertir todas las observaciones en puntuaciones Z. La media de la cantidad de exposiciones es 4,50, con un desvío estándar de 2,29. Por lo tanto, la primera observación, que es igual a 1, se ubica 3,5 unidades por debajo de la media, lo que implica 1,53 desvíos estándares debajo de la media, o sea $Z = -1,53$. Utilizando el mismo procedimiento para todas las otras observaciones se obtienen las puntuaciones Z que aparecen en las columnas correspondientes de la tabla 3-4. (La tabla *no* indica los pasos para el cómputo del desvío y del desvío cuadrático utilizados para calcular el desvío estándar).

b) Calcular el producto cruzado de las puntuaciones Z de cada persona. Por ejemplo, el primer producto cruzado es −1,53 por −0,74, lo que da un resultado de +1,13. Todos los productos cruzados aparecen en la columna ubicada a la derecha en la tabla 3-4.

c) Sumar los productos cruzados de las puntuaciones Z. El total es 10,80.

d) Dividir el resultado por la cantidad de personas. El resultado de dividir la suma de los productos cruzados de puntuaciones Z, 10,80, por la cantidad de personas, 16, es 0,68, que es el coeficiente de correlación. Es decir, $r = 0,68$.

5. Controlar el signo y el tamaño del coeficiente de correlación calculado comparándolo con la estimación realizada a partir del diagrama de dispersión. El resultado calculado de +0,68 es, como esperábamos, una correlación lineal positiva bastante marcada.

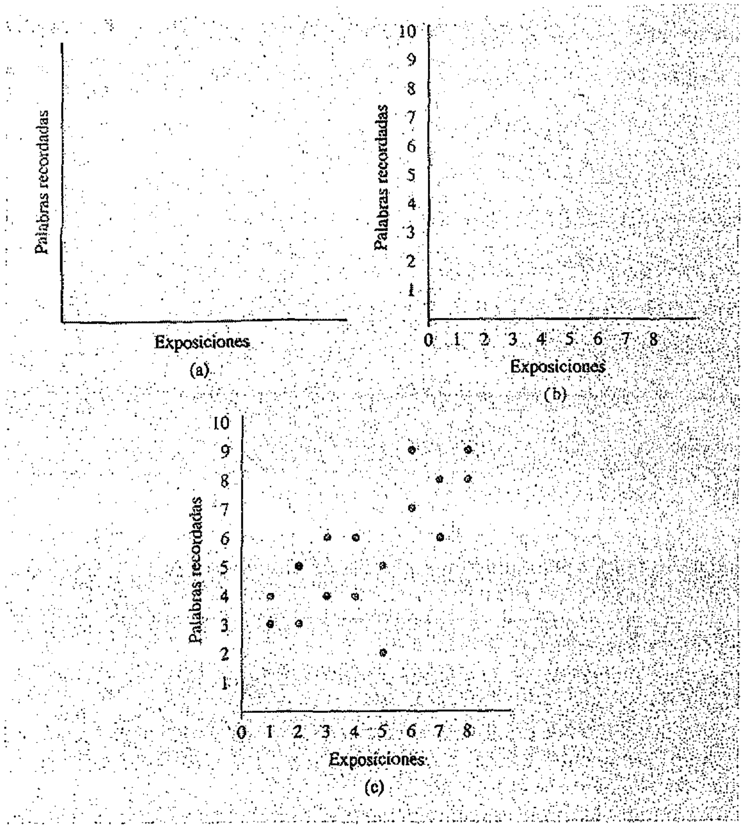


Figura 3-11. Pasos que se deben seguir para confeccionar un diagrama de dispersión según *los datos de la* tabla 3-3. (a) Establecer los ejes, la variable independiente (cantidad de exposiciones) en el eje horizontal, la variable dependiente (cantidad de palabras recordadas) en el eje vertical; (b) determinar la serie de valores, y marcarlos en los ejes; (c) ubicar un punto por cada par de observaciones de cada uno de los 16 participantes (datos ficticios).

Otro Ejemplo

Supongamos que una psicóloga educacional averiguó la cantidad promedio de alumnos por clase y los promedios de calificaciones en las pruebas de nivel de cinco escuelas primarias de determinado distrito escolar. La tabla 3-5 muestra los datos. La pregunta formulada por la psicóloga es: *¿Cuál es la relación entre estas dos variables?*

1. Construir un diagrama de dispersión.
   a) Dibujar los ejes y determinar en cuál se ubica cada variable. Dado que resulta razonable pensar que la cantidad de alumnos por clase afecta las calificaciones en las pruebas de nivel, y no al revés, podemos trazar en la parte inferior el eje correspondiente a la cantidad de alumnos por clase.

Tabla 3-4.
Cálculo del coeficiente de correlación del efecto producido por el número de exposiciones en la cantidad de palabras recordadas (datos ficticios).

Identificación del participante	Cantidad de esposiciones (variable independiente)		Cantidad de palabras recordadas (variable dependiente)		Producto cruzado de puntuaciones Z
	X	Z_X	Y	Z_Y	$Z_X Z_Y$
1	1	–1,53	4	–0,74	1,13
2	1	–1,53	3	–1,21	1,85
3	2	–1,09	3	–1,21	1,32
4	2	–1,09	5	–0,26	0,28
5	3	–0,65	6	0,21	–0,14
6	3	–0,65	4	–0,74	0,48
7	4	–0,22	4	–0,74	0,16
8	4	–0,22	6	0,21	–0,05
9	5	0,22	5	–0,26	–0,06
10	5	0,22	7	0,68	0,15
11	6	0,65	2	–1,68	–1,09
12	6	0,65	9	1,62	1,05
13	7	1,09	6	0,21	0,23
14	7	1,09	8	1,15	1,25
15	8	1,53	9	1,62	2,48
16	8	1,53	8	1,15	1,76
Σ:	72		89		10,80
M:	4,50		5,56		r = 0,68
$SD = \sqrt{84/16} =$	2,29		$\sqrt{72/16} =$ 2,12		

b) Determinar la serie de valores que se van a utilizar para cada variable y marcarla en los ejes. Presumiremos que las calificaciones en las pruebas de nivel van de 0 a 100. La cantidad de alumnos por clase debe ser por lo menos de 1 (y seguramente la política de la junta escolar exige que sean más). No conocíamos el máximo, así que supusimos que podía ser 50.

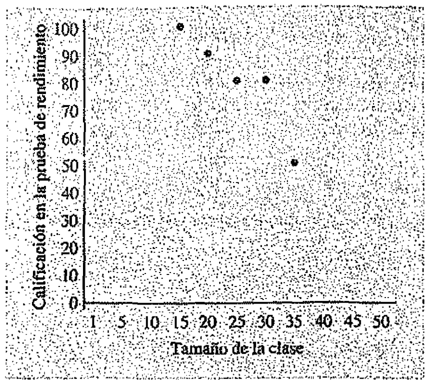
c) Marcar un punto por cada par de observaciones obtenidas de las personas (en este caso, de las escuelas). La figura 3-12 muestra el diagrama de dispersión completo.

2. Determinar si el diagrama muestra claramente una correlación curvilínea. En términos generales, la correlación parece mantener un patrón lineal (aunque con tan pocos puntos es difícil de decir).

Tabla 3-5.
Promedios de cantidad de alumnos por clase y de calificaciones en las pruebas de nivel en cinco escuelas primarias (datos ficticios).

Escuela primaria	Tamaño de la clase	Calificación en la prueba de rendimiento
Main Street	25	80
Casat	14	98
Harland	33	50
Shady Grove	28	82
Jefferson	20	90

Figura 3-12. Último paso en la confección de un diagrama de dispersión con la información contenida en la tabla 3-5: se ha dibujado un punto por cada par de observaciones de las cinco escuelas (datos ficticios).



3. Estimar la dirección y el grado de correlación lineal. Los puntos tienen una dirección marcada hacia abajo y hacia la derecha, indicando una fuerte correlación lineal negativa.
4. Calcular el coeficiente de correlación.
   a) Convertir todas las observaciones en puntuaciones Z. La media de la cantidad de alumnos por clase es 24 y el desvío estándar es 6,54. La puntuación Z de la cantidad de alumnos de la primera clase, 25, es igual a (25 – 24)/6,54 = 0,15. Todas las puntuaciones Z aparecen en la columna correspondiente de la tabla 3-6.
   b) Calcular los productos cruzados de las puntuaciones Z de cada persona (en este caso, de cada escuela). El primer producto cruzado es 0,15 x 0, que es igual a 0. El segundo es –1,53 x 1,10, que es igual a –1,68. Todos los productos cruzados de las puntuaciones Z aparecen en la columna de la derecha de la tabla 3-6.
   c) Sumar los productos cruzados de las puntuaciones Z. El total es –4,52.
   d) Dividir el total por la cantidad de personas (en este caso, escuelas). La suma (-4,52) dividida por 5 es igual a –0,90. Es decir, $r = -0,90$.
5. Controlar el signo y el tamaño del coeficiente de correlación calculado, comparándolo con la estimación realizada a partir del diagrama de dispersión. Un coeficiente de –0,90 concuerda perfectamente con la estimación original que indicaba una fuerte correlación lineal negativa.

Tabla 3-6.
Cálculo del coeficiente de correlación entre las cantidades promedio de alumnos por clase y de calificaciones en las pruebas de rendimiento en cinco escuelas primarias (datos ficticios).

Escuela	Tamaño de la clase		Calificación en la prueba de rendimiento		producto Cruzado
	X	Z_X	Y	Z_Y	$Z_X Z_Y$
Main Street	25	0,15	80	0,00	0,00
Casat	14	–1,53	98	1,10	–1,68
Harland	33	1,38	50	–1,84	–2,53
Shady Grove	28	0,61	82	0,12	0,08
Jefferson	20	–0,61	90	0,61	–0,38
Σ:	120		400		–40,52
M:	24		80		$r = -0,90$
	$SD = \sqrt{214/5} = 6,54$		$\sqrt{1,328/5} = 16,30$		

PRUEBA DE LA SIGNIFICACIÓN ESTADÍSTICA DEL COEFICIENTE DE CORRELACIÓN

Por sí mismo, el coeficiente de correlación es un estadístico descriptivo. Describe el grado y la dirección de la correlación lineal de determinado grupo de personas analizadas. Sin embargo, cuando realizamos una investigación en el campo de la psicología, por lo general estamos más interesados en una serie de observaciones en cuanto representan a una población mayor que no se ha analizado directamente. Por ejemplo, el psicólogo laboral entregó los cuestionarios sobre estrés sólo a cinco gerentes de la empresa, pero con la intención de considerarlos típicos representantes de los otros gerentes de esa misma empresa. (En la práctica se necesitaría un grupo con muchas más de cinco personas para lograr ese objetivo. Hemos utilizado cantidades pequeñas de personas en nuestros ejemplos para que sean más fáciles de comprender).

El problema, sin embargo, es que analizando sólo algunas de las personas es posible elegir por casualidad aquellas en las cuales los valores altos coinciden con los altos y los bajos con los bajos, aun cuando, habiendo estudiado a todas las personas, no hubiera existido correlación alguna. Decimos que una correlación es **significativa** si no resulta verosímil que hubiésemos podido *obtener una correlación de esa magnitud y si, en realidad, en el grupo completo no hubiera correlación alguna.* Específicamente, determinamos si esa verosimilitud es menor que algún bajo grado de probabilidad (p), como un 5% ó un 1%. Si esa verosimilitud es tan baja, decimos que la correlación es "estadísticamente significativa" con "$p < 0{,}05$" ó "$p < 0{,}01$."

El método y la lógica para determinar la **significación estadística** es el tema central de este libro a partir del capítulo 5. Estaríamos adelantando temas si intentáramos explicarlos ahora. De todos modos, para cuando hayamos completado los capítulos siguientes, la lógica y los detalles quedarán bien claros. (El apéndice II de este capítulo contiene la información necesaria para aplicar estos conocimientos a la correlación, pero en realidad no será muy útil hasta después de haber completado el capítulo 9). Sólo mencionamos el tema aquí para dar una idea general de lo que significa, en caso de que al leer alguna publicación científica que informe sobre coeficientes de correlación se haga referencia a la significación estadística "$p < 0{,}05$," o a alguna frase similar.

Cuadro 3-2. Correlación ilusoria: cuando estamos completamente seguros de que si es grande, es gordo... y estamos completamente equivocados.

El concepto de correlación no fue inventado en realidad por los especialistas en estadística. Es uno de los procesos mentales más básicos. Los primeros humanos deben haber pensado en términos de correlaciones todo el tiempo, al menos aquellos que sobrevivieron. "Cada vez que nieva, los animales que cazamos huyen. La nieve es sinónimo de ausencia de animales. Cuando vuelva a nevar tendremos que seguir a los animales para no morir de hambre".

De hecho, la correlación es un proceso mental tan típicamente humano que pareceríamos tener una organización psicológica tal que nos lleva a encontrar un grado de correlación mayor que el que en realidad existe, como ocurría con los aztecas, quienes pensaban que las buenas cosechas estaban correlacionadas con los sacrificios humanos (¡esperemos que hayan estado equivocados!), o como los siguientes ejemplos del proceso denominado **correlación ilusoria**, tomados de la psicología social.

Correlación ilusoria es el término que define la sobrestimación de la intensidad con la que se relacionan dos variables (también ha tenido otros significados especiales en el pasado). Sin duda, ya deben haber surgido en la mente del lector algunas correlaciones ilusorias étnicas, raciales, sexuales o relacionadas con la edad, realmente nocivas. Una fuente de correlación ilusoria es la tendencia a relacionar dos hechos poco frecuentes, y por lo tanto, fáciles de recordar. Supongamos que el grupo B es menos numeroso que el grupo A, y que se sabe que un tercio de las personas de ambos grupos comete, esporádicamente, algunas acciones indeseables. En este tipo de situaciones, las investigaciones demuestran que el grupo B, a cuyos miembros se los ve con menor frecuencia, será en efecto culpado por muchas más acciones no deseadas que el grupo A. Las cosas ocurren de ese modo aun cuando existan mayores chances de que determinada acción haya sido cometida por un integrante del grupo A, dado que tiene más miembros. El problema es que los hechos poco frecuentes se unen en la memoria. Ser miembro del grupo menos numeroso, y los comportamientos poco frecuentes, forman una correlación ilusoria. Una consecuencia obvia es que recordamos cualquier acto fuera de lo común, llevado a cabo por un miembro de una minoría, mucho mejor de lo que recordamos cualquier acto fuera de lo común realizado por un miembro de un grupo mayoritario.

La correlación ilusoria que nace como resultado de la "distinción de eventos apareados" (la vinculación mental de dos hechos poco usuales) puede tener lugar porque al encontrarnos por primera vez con experiencias diferentes pensamos más en ellas, procesándolas más profundamente, de modo que luego son más fáciles de recordar (Johnson & Mullen, 1994). Si nos encontramos, por ejemplo, con miembros de una minoría que no vemos con frecuencia, o con actos negativos que rara vez presenciamos o escuchamos, sin duda dedicaremos un tiempo a pensar en ellos. (Si los relacionamos en un par, los analizamos en conjunto y vuelven a nuestra memoria con mayor rapidez). Parecería que también pudiera ocurrir que continuemos procesando información acerca de grupos, personas y sus comportamientos, sin tener conciencia de esos procesos. En algún punto del proceso, o en momento de elaborar un juicio, asociamos más de lo debido a los grupos o personas que no encontramos con frecuencia con los comportamientos poco usuales (negativos) (McConnell, Sherman, & Hamilton, 1994).

De todos modos, la mayoría de las correlaciones ilusorias se producen como resultado de los prejuicios. Los prejuicios son teorías implícitas y erróneas que hemos incorporado. Por ejemplo, consideramos que hemos hallado una mayor confirmación de la relación entre dos características sociales de la que en realidad se desprende de lo observado: habilidad para conducir y determinada edad; nivel académico y grupo étnico específico; determinada forma de hablar, vestirse o comportamiento social y residencia en alguna región del país. Un ejemplo muy interesante es que la mayoría de los empresarios creen que el nivel de satisfacción laboral y el nivel de desempeño están estrechamente ligados cuando, en realidad, la correlación es bastante baja. Muchas personas a quienes no les gusta su trabajo pueden de todos modos desempeñarse a la perfección, mientras que otras muy entusiasmadas con su ocupación realizan su trabajo con desgano.

Lo importante es que la próxima vez que el lector se pregunte por qué está esforzándose tanto por aprender estadística, puede resultarle útil considerar que ese esfuerzo constituye una búsqueda destinada a lograr que sus procesos mentales sean más justos. Por ese motivo, volvemos a afirmar que la estadística puede transformarse en una empresa romántica, en tanto representa un modo de vencer errores malvados con la pureza de los números, de someter los prejuicios profanos con la honestidad de los datos.

Referencias: Hamilton (1981); Hamilton y Gifford (1976); Johnson and Mullen (1994).

CUESTIONES RELACIONADAS CON LA INTERPRETACIÓN DEL COEFICIENTE DE CORRELACIÓN

Un coeficiente de correlación describe la dirección y el grado de la correlación lineal entre dos variables. Sin embargo, al interpretar un coeficiente de correlación deben tomarse en cuenta algunas precauciones sutiles.

Causalidad y correlación

Si dos variables presentan una correlación lineal significativa, normalmente suponemos que existe algo que las correlaciona. Sin embargo, **la dirección de causalidad** (justamente, qué es la causa de qué) no puede determinarse solamente a partir de la correlación. En toda correlación entre dos variables X e Y, existen tres posibles direcciones de causalidad: X podría ser la causa de Y, Y la de Y, o algún tercer factor podría ser la causa de ambas, X e Y. También es posible (y a menudo probable) que exista más de una dirección de causalidad.

Tomemos el ejemplo del estrés de los gerentes. El estudio comenzó con la noción implícita de que supervisar un mayor número de personas (X) causa un aumento del nivel de estrés (Y). El resultado del estudio fue una marcada correlación positiva entre X e Y, que ciertamente coincide con la idea de que X es la causa de Y. Sin embargo, también coincide de la *misma forma* con la idea de que Y es la causa de X. (Tal vez los gerentes que parecen sufrir de estrés sean considerados muy trabajadores y ese sea el motivo por el cual sus superiores asignen mayor cantidad de personas a su cargo). También es posible que la correlación sea el resultado de algún tercer factor que cause que X e Y se desarrollen de manera conjunta. Por ejemplo, algunos sectores de la fábrica podrían necesitar más personal y también generar más estrés. Es decir, determinado sector de la fábrica causa estrés y requiere de muchos empleados para supervisar.

Existe bastante **confusión** acerca de este asunto de la correlación y la causalidad. El tema se complica al existir dos usos de la palabra **correlación**. Algunas veces se utiliza para describir un procedimiento estadístico (como lo hemos hecho en este capítulo), y otras veces se utiliza para describir un tipo de diseño de investigación en el que se miden dos variables en un grupo de personas, sin realizar una asignación aleatoria de sujetos a determinados valores de una de las variables (*véase* el apéndice A). Comúnmente, los **diseños de investigación** correlacionales son analizados estadísticamente utilizando el coeficiente de correlación, y los diseños de investigación experimentales se analizan utilizando procedimientos que veremos en los capítulos 9 al 13.

Sin embargo, existen excepciones. En este *mismo capítulo utilizamos un ejemplo* en el que los participantes eran asignados al azar en determinada cantidad de exposiciones y luego se medía la cantidad de palabras recordadas. A partir de los datos obtenidos, calculamos un coeficiente de correlación. No obstante, en el estudio no se utilizó un diseño de investigación correlacional; fue un verdadero experimento, ya que los participantes eran asignados al azar a diferentes valores de la variable independiente. Por sí mismo, el coeficiente de correlación que calculamos no nos indicó nada acerca de la causalidad. Aun así, quedó claro, por el diseño de investigación, que la única dirección causal posible es que la cantidad de exposiciones haya causado la diferencia en la cantidad recordada.

El coeficiente de correlación y la reducción proporcional de error

Un coeficiente de correlación indica la fuerza o el grado de una relación lineal; mayores valores de r (valores alejados de 0) indican un mayor grado de correlación. Es decir, una r de 0,4 significa que existe una correlación lineal más fuerte que una r de 0,2. Sin embargo, la mayoría de los investigadores sostendrían que una r de 0,4 no es el doble de fuerte que una r de 0,2. Para comparar correlaciones entre sí, la medida utilizada por la mayoría de los investigadores es r^2. A esto se lo denomina, por razones que veremos en el capítulo 4, **reducción proporcional del error** (y también **proporción de varianza explicada**).

Una correlación de 0,2 es equivalente a una r^2 de 0,04, y una correlación de 0,4 es equivalente a una r^2 de 0,16. *Por lo tanto, ¡una correlación de 0,4 implica* en realidad una relación cuatro veces más fuerte que una de 0,2!

Restricción del rango

Supongamos que un psicológo especializado en educación está interesado en la relación entre el grado que cursa un niño y los conocimientos sobre geografía. Si el investigador estudiara el rango completo de grados escolares, los resultados podrían ser como los que aparecen en el diagrama de dispersión de la figura 3-13a. Es decir, el investigador podría encontrar una fuerte correlación positiva. Pero supongamos que el investigador hubiera estudiado sólo los tres primeros grados (en

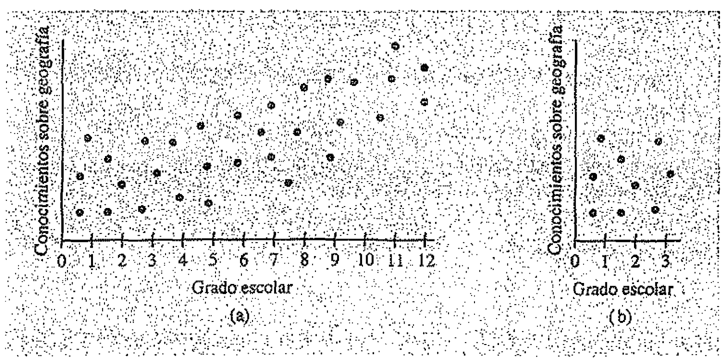


Figura 3-13. Ejemplo de restricción del rango comparando dos diagramas de dispersión: (a) cuando se muestra la serie de valores completa (grado escolar y conocimientos sobre geografía) y (b) cuando se restringe la serie de valores (a los primeros tres grados).

los que se enseña poca geografía). El diagrama de dispersión (*véase* figura 3-13b) reflejaría muy poca, o casi ninguna correlación (la correlación calculada sería cercana a 0) y, sin embargo, el investigador estaría incurriendo en error si llegara a la conclusión de que el grado no está relacionado con los conocimientos sobre geografía en ninguno de los grados escolares.

El problema en este caso es que la correlación está basada en una serie de observaciones que incluyen sólo un rango limitado de los valores posibles de una de las variables. (En este ejemplo existe un rango limitado de grados escolares). Es erróneo pensar en la correlación como si se aplicara a todo el rango de valores que podría tener la variable. Esta situación se denomina **restricción del rango**.

Es fácil cometer estos errores al interpretar correlaciones, las cuales incluso aparecen ocasionalmente en publicaciones científicas y se oyen con frecuencia aún mayor en discusiones informales sobre resultados de investigaciones. Por ejemplo, en el área de los negocios, a veces se intenta determinar si las pruebas de aptitud laboral reflejan lo exitosas que resultan ser en sus funciones las personas contratadas. Por lo general, la relación es baja, porque no se tiene en cuenta que se contratan sólo a las personas que tuvieron buenos resultados en las pruebas. Los estudios que miden el éxito en el empleo incluyen sólo el subgrupo que presenta los registros altos. La figura 3-14 grafica este ejemplo.

Falta de confiabilidad de la medición

Como hemos señalado, puede considerarse que el coeficiente de correlación describe la proximidad entre los puntos y una línea recta en el diagrama de dispersión. Sin embargo, una de las razones por las que los puntos pueden no estar ubicados cerca de la línea es el error casual en la medición; por ejemplo, un cuestionario que incluye algunos ítems ambiguos. Con frecuencia, en psicología, las mediciones no son perfectamente precisas o "confiables" (veremos este concepto en el capítulo 17 y en el apéndice A). El resultado es que una correlación calculada entre dos variables, tales como intimidad e idealización (para utilizar nuestro ejemplo anterior) resulta menor de lo que sería si tuviéramos medidas perfectas de estas variables.

La reducción en una correlación, debido a la falta de confiabilidad de las medidas, se denomina **atenuación**. Los libros de estadística más avanzada y los textos sobre medición psicológica describen fórmulas para la **corrección por atenuación**, que suponen que puede determinarse el grado de confiabilidad de las medidas, cuestión que no siempre es posible. En algunas publica-

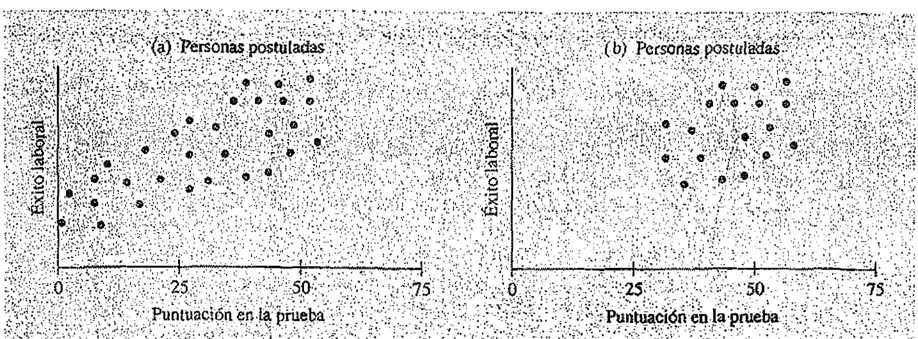


Figura 3-14. Ejemplo del efecto causado en la correlación por la restricción del rango.

ciones científicas podremos leer que la correlación ha sido "desatenuada" o que la "atenuación ha sido corregida". Significa que el investigador realizó ciertos cálculos para aumentar las correlaciones al nivel que se estimó que tendrían si se hubieran aplicado mediciones perfectas de las variables que se correlacionan.

Los procedimientos para corregir la atenuación superan el alcance de este libro. Aun así, es conveniente recordar la regla general que se aplica cuando las medidas son poco confiables, como ocurre en el caso de algunos cuestionarios, procedimientos observacionales, o en las observaciones psicológicas en general. Las correlaciones que aparecen en publicaciones que emplean medidas poco confiables pueden subestimar sustancialmente la verdadera correlación entre las variables que estas medidas imperfectas pretenden revelar.

CONTROVERSIAS Y DESARROLLOS RECIENTES: ¿QUÉ ES UNA GRAN CORRELACIÓN?

Con respecto al coeficiente de correlación existe en la actualidad una controversia sobre la definición de una "gran" *r*. Tradicionalmente, en psicología se consideraba una gran correlación aquélla que era igual o superior a 0,50; moderada, aquélla de aproximadamente 0,30, y pequeña, aquélla cercana a 0,10 (Cohen, 1988). De hecho, en psicología es raro obtener correlaciones mayores a 0,40. Aun cuando estemos seguros de que *X* es la causa de *Y*, seguramente no será la **única** causa. Si bien la intimidad provoca idealización, es sólo uno de los muchos factores que la causan. Todos los otros factores no forman parte de nuestra correlación. Ninguna correlación podría reflejar la historia completa. Las correlaciones pequeñas son también causadas por la poca confiabilidad de muchas medidas psicológicas.

Es tradicional advertir que una correlación pequeña no es muy importante aun cuando sea estadísticamente significativa. (Como veremos más adelante, una correlación pequeña puede ser estadísticamente significativa si el estudio incluye una gran cantidad de participantes). Después de todo, una correlación de 0,10 equivale sólo a un 1% de reducción del error.

Más aún, incluso psicólogos experimentados dedicados a la investigación, tienden a sobrestimar el grado de asociación que representa un coeficiente de correlación. Michael Oakes (1982), en la Universidad de Sussex, dio a 30 psicólogos dedicados a la investigación las dos columnas de datos que muestra la tabla 3-7. Luego les pidió que estimaran *r* (sin realizar ningún cálculo). ¿Qué

Tablea3-7.
Tabla presentada a 30 psicólogos para estimar *r*.

X	Y
1	1
2	10
3	2
4	9
5	5
6	4
7	6
8	3
9	11
10	8
11	7
12	12

Fuente: Oakes (1982).

opina el lector? La intuición de los investigadores británicos (que como grupo están, al menos, tan bien capacitados en estadística como los psicólogos de cualquier lugar del mundo) indicaba desde –0,20 a +0,60, con una media de 0,24. Si el lector lo desea puede calcular la verdadera correlación. ¡Es de 0,50! Es decir, que en forma abstracta los psicólogos dan a una correlación de 0,50 un grado mucho más alto de correlación del que le otorgan cuando observan los datos reales (datos que, aun con una $r = 0{,}50$, sólo se **veían** como de 0,24).

Oakes dió a otro grupo de treinta investigadores sólo la columna de *X*, y les pidió que completaran la columna de *Y* con números tales que reflejaran una correlación de 0,50 (nuevamente, sólo utilizando su intuición y sin realizar ningún cálculo). Cuando Oakes calculó las correlaciones reales que representaban los números indicados por los investigadores, el promedio resultó ser de 0,68. En otras palabras, incluso los investigadores experimentados consideran que un coeficiente de correlación representa un grado de asociación mayor de lo que en realidad implica.

Por el contrario, otros psicólogos sostienen que las pequeñas correlaciones pueden ser muy importantes teóricamente, las cuales tienen mayor inferencia práctica por el hecho de que los pequeños efectos pueden acumularse a lo largo del tiempo (Prentice & Miller, 1992).

Para demostrar la importancia práctica de pequeñas correlaciones, Rosnow y Rosenthal (1989b) ofrecen un ejemplo de un estudio actualmente famoso relacionado con el hecho de que algunos médicos tomaran o no aspirinas a diario, y la relación de esa ingesta con los ataques cardíacos (Comisión Directiva del Grupo Médico de Investigación sobre Estudios Sanitarios [Steering Committee of the Physicians Health Study Research Group], 1988). Los resultados demostraron que el hecho de ingerir aspirina estaba correlacionado en un –0,034 con los ataques cardíacos.[1] Es decir, produce aproximadamente un 0,1% de reducción proporcional de error. Sin embargo, consideremos la parte superior de la tabla 3-8 (en la que se describen los ataques cardíacos con IM (Infarto demiocardio). La correlación de "sólo 3,4" significaba que entre los más de 20.000 médicos que estaban incluidos en el estudio, existían 72 ataques cardíacos más en el grupo que no tomaba aspirina. (De hecho, la parte inferior de la tabla indica que también existían 13 muertes más por ataques cardíacos en el grupo que no tomaba aspirina).

La parte central de la tabla 3-8 es especialmente interesante desde el punto de vista de la estadística. En esa parte de la tabla, los porcentajes correspondientes a cada grupo referido a los ataques cardíacos (ausencia de IM y presencia de IM) están divididos entre el porcentaje que tomaba y el que no tomaba aspirinas (el grupo placebo). Observemos que la diferencia en porcentajes en cualquiera de los grupos es de exactamente +3,4% ó –3,4%. (Por ejemplo, dentro del grupo que había sufrido ataques cardíacos, el 48,3% que tomaba aspirinas menos el 51,7% del grupo placebo equivale a –3,4%). Lo importante aquí es que 3,4% es exactamente igual al coeficiente de correlación (0,034).

No es una coincidencia. En este tipo de tablas, que se denominan **del tamaño del efecto a una exposición dicotómica**, la diferencia de los porcentajes siempre resultará exactamente igual al coeficiente de correlación. Esto facilita sorprendentemente la comprensión de las correlaciones en cualquier situación en la que una de las variables representa el éxito y el fracaso (como no sufrir un ataque cardíaco o sufrirlo), y la otra variable representa la clase de tratamiento recibido (como aspirina o placebo). En este tipo de situaciones, la correlación marca la diferencia en porcentajes entre el éxito y el fracaso en relación con el tratamiento recibido.

[1] Para calcular la correlación entre tener un ataque cardíaco y tomar aspirinas, tendríamos que convertir las dos variables en números. Por ejemplo, podríamos representar el hecho de tener un ataque cardíaco con 1, y no tenerlo con 0; de forma similar, podríamos considerar que estar en el grupo que consume aspirinas es igual a 1, y estar en el grupo placebo igual a 0. No tiene importancia cuál de los dos números utilicemos para cada uno de los dos valores de cada variable. Cualesquiera sean los dos números utilizados, el resultado será el mismo después de convertirlos en puntuaciones Z. La única diferencia que puede surgir en relación con los números utilizados es que, según a qué valor se aplique el número mayor, esto determinará que la correlación sea positiva o negativa.

Tabla 3-8.
Efectos de la aspirina en los ataques cardíacos.

Condición	Ausencia de IM	Presencia de IM
Aspirina	10.933	104
Placebo	10.845	189
	Tamaño del efecto a una exposición dicotómica	
Aspirina	51,7	48,3
Placebo	48,3	51,7
Total	100,0	100,0
	IM no fatal	**IM fatal**
Aspirina	99	5
Placebo	171	18

Nota: IM = infarto de miocardio. **Fuente:** Comisión Directiva del Grupo Médico de Investigación Sobre Estudios Sanitarios [Steering Committee of the Physicians Health Study Research Group] (1988).

Ciertamente, un 3,4% de diferencia entre sufrir o no ataques cardíacos es un dato interesante, aunque es sólo una pequeña parte de lo que afecta a las personas que sufren ataques cardíacos. No deja de ser cierto que el 99,9% de la variación, en cuanto a que la gente sufra o no ataques cardíacos, se debe a otros factores (dieta, ejercicio, herencia, etc.). Más aún, algunos estadísticos (por ejemplo Strahan, 1991; Thompson & Schumacker, 1997) sostienen que este tamaño del efecto a la exposición dicotómica distorsiona la situación real, excepto cuando la frecuencia de los valores de ambas variables están en la relación 50-50. (En este caso, la relación entre aspirina y placebo es cercana al 50-50, pero con respecto a presencia y ausencia de IM la relación está muy lejos de ser 50-50).

Otra discusión sobre la importancia de las pequeñas correlaciones está basada en los métodos de investigación. Prentice y Miller (1992) explican:

> Demostrar que un efecto perdura aun bajo las circunstancias más inverosímiles puede ser tan impresionante (o en algunos casos, tal vez más impresionante) que demostrar que el efecto contribuye en una gran parte a la varianza (p. 163).

Algunos de los ejemplos de estos autores se refieren a estudios que demuestran una correlación entre el sentimiento de atracción y las sentencias de culpabilidad o inocencia emitidas en juicios (por ejemplo Sigall & Ostrove, 1975). Lo importante es que "las sentencias legales no deberían ser afectadas por factores tan accidentales como el sentimiento de atracción". Por lo tanto, si existen estudios que demuestran que la atracción está relacionada con las sentencias, aunque más no sea levemente, entonces nos convenceríamos de la importancia que podría tener la atracción por su influencia en las opiniones sociales en general.

COEFICIENTES DE CORRELACIÓN SEGÚN SE DESCRIBEN EN LAS PUBLICACIONES CIENTÍFICAS

Los coeficientes de correlación aparecen en las publicaciones científicas tanto en el texto como en las tablas (algunas veces también se hace referencia al "nivel de significación", como por ejemplo, $p < 0{,}05$). El resultado del estudio con el que comenzamos el capítulo se describiría en el texto de la siguiente manera: "Existía una fuerte correlación positiva entre la idealización y la intimidad, $r = 0{,}74$."

Las tablas de correlaciones son muy comunes cuando se trabaja con varias variables. Usualmente, se diseña la tabla de modo que cada variable aparezca tanto en la parte superior como en el margen izquierdo, y la correlación entre cada par de variables se indica dentro de la tabla, a la que se denomina **matriz de correlación**.

La tabla 3-9 proviene de una publicación que describe los resultados de un gran estudio europeo sobre dietas alimenticias saludables (Wardle et al., 1997). En este caso en particular, los autores incluyen las matrices de correlación separadas una para mujeres y otra para hombres. El ejemplo que presentamos aquí ilustra varias características típicas del modo en que se diseñan las matrices de correlación. En primer lugar, podemos observar que no se indica la correlación de una variable consigo misma. En este ejemplo, el espacio se completa con un guión; con frecuencia sólo se deja el espacio libre. Podemos observar también que sólo la mitad superior de cada matriz está completa. Esto ocurre porque completar la otra mitad sería repetitivo; por ejemplo, la correlación entre evitar las grasas con consumir fibras es la misma que la correlación entre consumir fibras y evitar las grasas (en algunos casos se completa la mitad inferior y se deja en blanco la mitad superior). Existe otra forma de resumir la información ahorrando espacio en la página: los nombres de las variables aparecen sólo en el costado de la tabla; en la parte superior sólo se escriben los números correspondientes a cada una de ellas. Finalmente, observamos que los coeficientes significativos están indicados con un asterisco, y se incluye una nota al pie que explica el nivel de probabilidad que representa el asterisco.

Observando el ejemplo, podemos ver, entre otras cosas, que las correlaciones entre evitar grasas y consumir fibras son bastante altas (0,42 en el caso de las mujeres y 0,41 en el caso de los hombres). También es interesante el hecho de que exista muy poca o ninguna correlación entre consumir frutas diariamente y limitar las carnes rojas (0,00 para las mujeres y 0,01 para los hombres).

Tabla 3-9.
Correlaciones de Pearson entre las costumbres alimenticias de hombres y de mujeres.

Sub-escala	1	2	3	4	5
			Mujeres ($n = 9{,}182$)		
1. Evitar las grasas	—	0,42*	0,16*	0,14*	0,11*
2. Consumir fibras		—	0,15*	0,12*	0,09*
3. Comer frutas diariamente			—	0,05*	0,00
4. Limitar las carnes rojas				—	0,12*
5. Limitar la sal					—
			Hombres ($n = 7.304$)		
1. Evitar las grasas	—	0,41*	0,13*	0,12*	0,10*
2. Comer fibras		—	0,13*	0,11*	0,08*
3. Comer frutas diariamente			—	0,02*	0,01*
4. Limitar las carnes rojas				—	0,07*
5. Limitar la sal					—

*$p < 0{,}001$.

Fuente: Wardle, J., et al. (1997), tab. 2. "Prácticas alimenticias saludables de alumnos europeos. "*Psicología sanitaria*", 16, 443-450. Copyright, 1997, por la Asociación Americana de Psicología [American Psychological Association]. Reimpreso con autorización.

RESUMEN

Un diagrama de dispersión muestra la relación entre dos variables. En el eje horizontal se ubican los valores de la variable independiente o predictora, ordenados de menor a mayor. En el eje vertical se ubican los valores de la variable dependiente, ordenados de menor a mayor. Cada par de valores correspondientes a un individuo se marca con un punto.

Cuando en términos generales los puntos del diagrama de dispersión siguen una línea recta, hablamos de una correlación lineal. En una correlación lineal positiva, la recta va hacia arriba y hacia la derecha (es decir, los valores bajos coinciden con los bajos y los altos con los altos). En una correlación lineal negativa, la recta va hacia abajo y hacia la derecha (es decir, los valores bajos coinciden con los altos y los altos con los bajos). En una correlación curvilínea, los puntos siguen un patrón distinto de una simple línea recta. Existe correlación nula cuando los puntos no siguen ningún tipo de patrón sistemático.

El coeficiente de correlación (r) indica el grado de correlación lineal. Es el promedio de los *productos cruzados de puntuaciones Z*. Cuando existe una fuerte correlación lineal positiva, el coeficiente de correlación es altamente positivo debido a que las puntuaciones Z positivas se multiplican por positivas y las puntuaciones Z negativas por negativas. Cuando existe una fuerte correlación lineal negativa, el coeficiente de correlación es altamente negativo debido a que las puntuaciones Z positivas se multiplican por negativas y las puntuaciones Z negativas por positivas. Cuando no existe correlación lineal, el coeficiente de correlación es 0, debido a que las puntuaciones Z positivas son multiplicadas a veces por puntuaciones Z positivas, y otras por puntuaciones Z negativas, mientras que las puntuaciones Z negativas son multiplicadas a veces por puntuaciones Z negativas, y otras por puntuaciones Z positivas. Por lo tanto, los productos cruzados positivos y negativos se cancelan entre sí.

El máximo valor positivo posible de r es +1, $r = +1$, y ocurre cuando existe una correlación lineal positiva perfecta. El máximo valor negativo posible de r es -1, $r = -1$, y ocurre cuando existe una correlación lineal negativa perfecta.

Una correlación generalmente está basada en valores observados de determinado grupo que pretende representar a un grupo más amplio. Cuando los resultados de los procedimientos estadísticos (que aprenderemos más adelante) no son coherentes con la idea de que la correlación en ese grupo más amplio es 0, decimos que la correlación es estadísticamente significativa.

Las comparaciones del grado de correlación lineal se consideran más precisas si se realizan con el cuadrado del coeficiente de correlación (r^2), llamado reducción proporcional del error.

La correlación no muestra la dirección de causalidad. Si dos variables, X e Y, están correlacionadas, esto podría ser porque X está causando Y, Y está causando X, o un tercer factor está causando X e Y.

Un coeficiente de correlación puede representar la verdadera correlación por debajo de su nivel verdadero si se basa en las observaciones de un grupo de estudio cuyo rango de valores es restringido, o cuyos valores se basan en medidas poco confiables.

Muchos psicólogos sostienen que el coeficiente de correlación es una sobrestimación de la importancia de la asociación entre dos variables. En efecto, los estudios realizados sugieren que los psicólogos tienden a considerar cualquier coeficiente de correlación en particular como representante de un mayor grado de asociación del que realmente existe. Sin embargo, las pequeñas correlaciones pueden tener importancia práctica (que puede ser demostrada a través del tamaño del efecto a una exposición dicotómica, el cual describe la relación entre dos variables con dos valores cada una, y examinando la tabla 2 x 2 resultante). Las pequeñas correlaciones también pueden ser muy efectivas para demostrar la importancia de una relación cuando un estudio demuestra que la correlación se mantiene aun bajo lo que parecerían condiciones poco probables.

Las publicaciones científicas generalmente presentan resultados correlacionales tanto en sus textos, con el valor r (y algunas veces con el nivel de significación), como en tablas especiales (matrices de correlación) que ilustran las correlaciones entre diversas variables.

Términos clave

- Corrección por atenuación.
- Correlación.
- Coeficiente de correlación (r).
- Matriz de correlación.
- Producto cruzado de puntuaciones Z.
- Correlación curvilínea.
- Grado de correlación.
- Variable dependiente.
- Dirección de causalidad.
- Variable independiente.
- Correlación lineal.
- Correlación negativa.
- Correlación nula.
- Correlación perfecta.
- Correlación positiva.
- Variable predictora.
- Reducción proporcional del error (r^2).
- Restricción de rango.
- Diagrama de dispersión.
- Significación estadística.

Ejercicios

Los ejercicios implican la realización de cálculos (con la ayuda de una calculadora).

La mayoría de los problemas estadísticos reales se resuelven por computadora, pero aunque exista la posibilidad de utilizarla, es conveniente realizar estos ejercicios manualmente para incorporar el método de trabajo.

Para adquirir práctica en la utilización de una computadora, para resolver problemas estadísticos, se puede utilizar la sección de computación de cada capítulo, publicada en la *Guía de estudio y libro de tareas de computación para el alumno [Student's Study Guide and Computer Workbook]* que acompaña este libro.

Todos los datos de esta sección son ficticios (a menos que se especifique lo contrario).

Las respuestas a los ejercicios de la serie I se encuentran al final del libro.

SERIE 1

Realice las siguientes tareas para los ejercicios 1 y 2: a) Confeccione un diagrama de dispersión con las puntuaciones originales; b) describa con palabras el patrón general de correlación, si existe; c) calcule el coeficiente de correlación; d) explique la lógica de lo que ha hecho, escribiendo como si se dirigiera a alguien que nunca ha asistido a un curso de estadística (pero que sí entiende qué es la media, el desvío estándar y las puntuaciones Z), y e) presente tres direcciones de causalidad lógicamente posibles, indicando en cada caso si se trata de una explicación razonable para la correlación según las variables involucradas (y por qué).

1. Un investigador estaba interesado en la relación entre el grado de empatía de los psicoterapeutas y el nivel de satisfacción de sus pacientes con la terapia. Como estudio piloto se analizaron cuatro parejas de terapeutas y pacientes. Estos son los resultados:

Número de pareja	Empatía del terapeuta	Satisfacción del paciente
1	70	4
2	94	5
3	36	2
4	48	1

2. Un instructor preguntó a cinco alumnos cuántas horas habían estudiado para un examen. A continuación se detalla la cantidad de horas de estudio y sus calificaciones.

Horas de estudio	Calificación en la prueba
0	52
10	95
6	83
8	71
6	64

3. En un estudio realizado a personas que recién se conocían, se midió el nivel de extraversión de uno de los integrantes de la pareja y el aprecio del otro integrante de la pareja por el primero. Estos son los resultados:

Extraversión de uno de los integrantes		Aprecio por ese integrante	
Puntuación original	*Puntuación Z*	*Puntuación original*	*Puntuación Z*
18	0,37	8	1,10
17	0,17	9	1,47
20	0,80	6	0,37
8	−1,72	1	−1,47
13	−0,67	7	0,74
24	1,63	1	−1,47
11	−1,09	3	−0,74
12	−0,88	5	0,00
18	0,38	7	0,74
21	1,00	3	−0,74

En este ejercicio damos las puntuaciones Z para ahorrar tiempo de cálculo. a) Construya un diagrama de dispersión de las puntuaciones originales; b) describa con palabras el patrón general de la asociación, si existe, y c) calcule el coeficiente de correlación.

4. Chapman, Hobfoll y Ritter (1997) entrevistaron dos veces durante el embarazo a 68 mujeres de una zona céntrica y superpoblada de una ciudad y a sus maridos (o novios); la primera vez, entre el tercer y sexto mes de embarazo, y la siguiente vez, entre el sexto y el noveno mes de embarazo. La tabla 3-10 muestra las correlaciones entre varias de las medidas. Lo más importante en esta tabla es la correlación entre lo que las mujeres informaban sobre su propio estrés, lo que los hombres informaban sobre el estrés de sus compañeras, la percepción de las mujeres sobre el apoyo brindado por sus parejas en la primera y en la segunda entrevista y el nivel de depresión de las mujeres en la primera y en la segunda entrevista.

Explique los resultados de las medidas como si estuviera escribiendo para una persona que nunca ha asistido a un curso de estadística. Específicamente, a) explique qué significa un coeficiente de correlación, utilizando una de las correlaciones como ejemplo; b) analice la tabla y luego comente los patrones de los resultados, indicando las variables que presentan una correlación relativamente fuerte y las que no, y c) comente las limitaciones que deben tenerse en cuenta al sacar conclusiones sobre causalidad sobre la base de esta información, utilizando como ejemplo una correlación específica (nombre al menos una dirección de causalidad alternativa posible y explique por qué esa alternativa es posible).

5. Para cada una de las siguientes situaciones, indique por qué el coeficiente de correlación podría ser una estimación distorsionada de la correlación real (y qué clase de distorsión esperaría):

a) Puntuaciones en dos cuestionarios de medición de personalidad están correlacionados.

b) La calidad de vida y la felicidad de un grupo de millonarios están correlacionadas.

6. La siguiente información ha sido preparada de forma tal que las series de datos B hasta D sean versiones levemente modificadas de la serie A. Confeccione diagramas de dispersión y calcule los coeficientes de correlación de cada serie de datos (sólo damos la *solución de las series* A y B).

Serie A		Serie B		Serie C		Serie D	
X	*Y*	*X*	*Y*	*X*	*Y*	*X*	*Y*
1	1	1	1	1	5	1	1
2	2	2	2	2	2	2	4
3	3	3	3	3	3	3	3
4	4	4	5	4	4	4	2
5	5	5	4	5	1	5	5

7. Un investigador está interesado en averiguar si un nuevo medicamento produce algún efecto en caso de resfrío. Ocho personas son analizadas: cuatro toman el medicamento y cuatro no (las que lo toman son calificadas con

un 1, las que no, con un 0) y luego se registra si se resfrían (calificación 1) o no (calificación 0). A continuación aparecen cuatro resultados posibles. Calcule el coeficiente de correlación en cada caso (sólo damos la solución para las posibilidades A y B).

Posibilidad A		Posibilidad B		Posibilidad C		Posibilidad D	
Toma Med.	*Se resfría*	*Toma Med.*	*Se resfría*	*Toma Med.*	*Se resfría*	*Toma Med.*	*Se resfría*
0	1	0	1	0	1	0	1
0	1	0	1	0	1	0	1
0	1	0	1	0	0	0	1
0	1	0	0	0	0	0	0
1	0	1	1	1	1	1	0
1	0	1	0	1	1	1	0
1	0	1	0	1	0	1	0
1	0	1	0	1	0	1	0

SERIE II

Realice lo siguiente en los ejercicios 1 y 2: a) construya un diagrama de dispersión de las puntuaciones originales; b) describa con palabras el patrón general de correlación, si existe; c) calcule el coeficiente de correlación; d) explique la lógica de lo que ha hecho, escribiendo como si estuviera haciéndolo para alguien que nunca asistió a un curso de estadística (pero que sí comprende qué es la media, el desvío estándar y las puntuaciones Z), y e) indique tres direcciones de causalidad lógicamente posibles, explicando en cada caso si es una dirección razonable de la correlación según las variables involucradas (y ¿por qué?).

1. Se entrega a cuatro individuos una prueba de habilidad manual (los valores altos significan mayor habilidad) y una prueba de ansiedad (los valores altos implican mayor ansiedad). Los valores observados de los cuatro individuos son los siguientes:

Persona	Habilidad	Ansiedad
1	1	10
2	1	8
3	2	4
4	4	-2

2. Se controla de cerca a cuatro niños pequeños durante un periodo de varias semanas para medir qué cantidad de programas de televisión violenta miran y la medida de su comportamiento violento hacia sus compañeros de juego. Los resultados fueron los siguientes:

Tabla 3-10.
Correlaciones de orden cero de las variables del estudio.

Variable	1	2	3	4	5	6	7	8	9	10
1. Estrés informado por mujeres	—									
2. Estrés de mujeres informado por hombres	0,17	—								
3. Apoyo de pareja 1	–0,28*	–0,18	—							
4. Apoyo de pareja 2	–0,27*	–0,18	0,44***	—						
5. Estado depresivo 1	0,23*	0,10	–0,34**	–0,17	—					
6. Estado depresivo 2	0,50***	0,14	–0,42***	–0,41***	0,55***	—				
7. Edad de las mujeres	0,06	0,16	0,04	–0,24*	–0,35*	–0,09	—			
8. Origen étnico de las mujeres	–0,19	–0,09	–0,16	–0,14	0,11	0,13	–0,02	—		
9. Estado civil de las mujeres	–0,18	0,01	0,12	0,24*	–0,04	–0,20	0,05	–0,34**	—	
10. Paridad	0,19	0,13	–0,11	–0,17	0,10	0,16	0,26*	0,31*	–0,12	—

*$p < 0{,}05$; **$p < 0{,}01$; ***$p < 0{,}001$.

Fuente: Chapman, H. A., Hobfoll, S. E., & Ritter, C. (1997), tab. 2. "El hecho de que el compañero subestime el estrés sufrido por ellas provoca angustia en las mujeres: estudio sobre mujeres embarazadas de zonas céntricas y superpobladas de la ciudad". *Periódico sobre Psicología Social y de Personalidad [Journal of Personality and Social Psychology]*, 73, 418-425. Copyright, 1997, por la Asociación Americana de Psicología [American Psychological Association]. Reimpreso con autorización.

Número de código de los niños	Cantidad semanal (horas) de TV violentas	Cantidad de acciones violentas o agresivas hacia compañeros
G3368	14	9
R8904	8	6
C9890	6	1
L8722	12	8

En los ejercicios 3 y 4, a) construya un diagrama de dispersión de las puntuaciones originales; b) describa con palabras el patrón general de correlación, si existe, y c) calcule el coeficiente de correlación. En los dos ejercicios damos las puntuaciones Z para ahorrarle tiempo.

3. Supongamos que el Museo de Louvre está interesado en la relación entre la antigüedad de una pintura y el interés del público en esa pintura. Durante una semana se controla la cantidad de personas que se detienen a observar a cada una de las 10 pinturas elegidas al azar. Los resultados son los siguientes:

Título de la pintura	Antigüedad aproximada (años)		Cantidad de personas que se detienen a observarla	
	X	Z_X	X	Z_Y
El Entierro	465	1,39	68	−0,69
Mys Mar Ste Catherine	515	1,71	71	−0,59
Las Bañistas	240	−0,09	123	1,19
El Toilette	107	−0,96	112	0,82
Retrato de Castiglione	376	0,80	48	−1,38
Carlos I de Inglaterra	355	0,67	84	−0,14
Crispin y Scapin	140	−0,75	66	−0,76
Desnudo al Sol	115	−0,91	148	2,05
El Balcón	122	−0,86	71	−0,59
El Circo	99	−1,01	91	0,10

4. Un maestro de escuela creyó notar que los alumnos que se vestían más prolijamente eran, en líneas generales, mejores estudiantes. Para probar esta idea, el maestro hizo que un amigo calificara a cada uno de los alumnos según su prolijidad en el vestir. A continuación detallamos las calificaciones por prolijidad, junto con las calificaciones de los alumnos en una prueba estandarizada de rendimiento escolar.

Niño	Calificación por prolilidad		Registros en prueba de nivel	
	X	Z_X	X	Z_Y
Janet	18	−0,52	60	−0,66
Gareth	24	1,43	58	−1,09
Grove	14	−1,82	70	1,47
Kevin	19	−0,20	58	−1,09
Joshua	20	0,13	66	0,62
Nicole	23	1,11	68	1,04
Susan	20	0,13	65	0,40
Drew	22	0,78	68	1,04
Marie	15	−1,50	56	−1,51
Chad	21	0,46	62	−0,23

5. Como parte de un estudio más amplio, Speed y Gangestad (1997) obtuvieron calificaciones y nominaciones sobre diversas características de 66 hombres de una fraternidad, otorgadas por sus compañeros de fraternidad. El siguiente párrafo fue tomado de la sección de resultados del estudio:

> La popularidad romántica de los hombres estaba significativamente correlacionada con varias características: mejor vestimenta (r = 0,48), mayor atractivo físico (r = 0,47), más sociabilidad (r = 0,47), más confianza en sí mismo (r = 0,44), mejor líder (r = 0,38), más divertido (r = 0,37), más satisfecho (r = 0,32) y más independiente (r = 0,28). Sin embargo, inesperadamente, el potencial de los hombres en relación con el éxito financiero no estaba significativamente correlacionado con su popularidad romántica (r = 0,10). (p. 931).

Explique los resultados como si estuviera escribiendo para una persona que nunca ha asistido a un curso de estadística. Específicamente, a) explique qué significa un coeficiente de correlación utilizando una de las correlaciones como ejemplo; b) explique qué significa "significativamente" y "no significativamente", en general, refiriéndose al menos a un ejemplo específico y c) especule sobre el significado del patrón de los resultados, teniendo en cuenta el tema de la dirección de causalidad.

6. Seleccione arbitrariamente ocho nombres personales completos, de ocho hojas diferentes de la guía telefónica. Confeccione un diagrama de dispersión y calcule el coeficiente de correlación entre la cantidad de letras en el primer nombre y en el apellido. Describa el resultado con palabras y sugiera una posible interpretación de sus resultados.

APÉNDICE I DEL CAPÍTULO: FÓRMULA DE CÁLCULO OPTATIVA DEL COEFICIENTE DE CORRELACIÓN

Los pasos para calcular un coeficiente de correlación pueden combinarse en una sola fórmula para realizar cálculos a mano (o con una calculadora) en un estudio con gran cantidad de participantes. Comúnmente, los pasos para calcular una correlación son: calcular a) las puntuaciones Z de cada puntuación original, b) los productos cruzados de las puntuaciones Z y c) el promedio de los productos cruzados de las puntuaciones Z. (El alumno habrá notado, al realizar los ejercicios, que calcular las puntuaciones Z es particularmente tedioso cuando se trabaja a mano, especialmente si primero es necesario calcular las medias y los desvíos estándares). Con un poco de manipulación algebraica, la fórmula puede transformarse en la que se indica a continuación. (Aunque parezca terrible, resulta realmente más sencilla para aplicar en un estudio real con grandes cantidades de participantes que si tuvieran que calcularse los resultados a mano).

$$r = \frac{N\Sigma XY - (\Sigma X)(\Sigma Y)}{\left[\sqrt{N\Sigma X^2 - (\Sigma X)^2}\right]\left[\sqrt{N\Sigma Y^2 - (\Sigma Y)^2}\right]} \qquad (3\text{-}2)$$

Cuando se utiliza este procedimiento resulta útil organizar los datos en un cuadro de cinco columnas, formado por las columnas X, X^2, Y, Y^2, y los productos cruzados de XY. Cabe destacar que no se mencionan las puntuaciones Z, y que los productos cruzados se calculan directamente sobre la base de puntuaciones originales. Además, tal como lo recordamos en el apéndice del capítulo 2, ΣX^2 se logra tomando cada valor X y elevándolo al cuadrado, y luego sumando estos cuadrados; por el contrario, $(\Sigma X)^2$ se logra sumando todos los valores X (sin elevar al cuadrado ninguno de ellos), y luego elevando el total al cuadrado.

La tabla 3-11 muestra el cálculo correspondiente al ejemplo del estrés de los gerentes utilizando esta fórmula. Compárela con la tabla 3-2.

APÉNDICE II DEL CAPÍTULO: PRUEBA DE HIPÓTESIS Y SU POTENCIA SOBRE EL COEFICIENTE DE CORRELACIÓN

Este material está dirigido a aquellos alumnos que ya han completado al menos hasta el capítulo 9 y ahora vuelven a estudiar este capítulo.

Significación de un coeficiente de correlación

La prueba de hipótesis de un coeficiente de correlación sigue el proceso usual de cinco pasos. Sin embargo, cabe destacar tres puntos importantes. Primero, la hipótesis nula establece, en líneas generales, que **la correlación en una población como la observada no es diferente de la de una población en la que la verdadera correlación es 0.** Segundo, si los supuestos (explicados en el siguiente párrafo) se cumplen, la distribución comparativa es una distribución t con grados de libertad iguales a la cantidad de participantes menos 2. Tercero, el estadístico de prueba que corresponde al de correlación en esa distribución t se calcula utilizando la siguiente fórmula:

$$t = \frac{(r)\left(\sqrt{N-2}\right)}{\sqrt{1-r^2}} \qquad (3\text{-}3)$$

Tabla 3-11.
Cómputos del coeficiente de correlación del estudio sobre el estrés de los gerentes, realizados con la fórmula de cálculo (datos ficticios).

	Empleados supervisados		Nivel de estrés		Productos cruzados
	X	X^2	Y	Y^2	XY
	6	36	7	49	42
	8	64	8	64	64
	3	9	1	1	3
	10	100	8	64	80
	8	64	6	36	48
Σ:	35	273	30	214	237

$$r = \frac{N\Sigma XY - (\Sigma X)(\Sigma Y)}{[\sqrt{N\Sigma X^2 - (\Sigma X)^2}]\ [\sqrt{N\Sigma Y^2 - (\Sigma Y)^2}]}$$

$$r = \frac{(5)(237) - (35)(30)}{[\sqrt{(5)(273) - (35)^2}]\ [\sqrt{(5)(214) - (30)^2}]}$$

$$r = \frac{1.185 - 1.050}{(\sqrt{1.365 - 1.225})\ (\sqrt{1.070 - 900})}$$

$$r = \frac{135}{(\sqrt{140})(\sqrt{170})} = \frac{135}{(11,83)(13,04)} = \frac{135}{154,26} = 0,88$$

Además, queremos destacar que las pruebas de significación de una correlación, como por ejemplo una prueba t, pueden ser de una o dos colas. Una prueba de una cola significa que el investigador ha predicho el signo (positivo o negativo) de la correlación[2].

Los supuestos de las pruebas de significación de un coeficiente de correlación son algo complejos. Comúnmente, ambas variables deberían estar normalmente distribuidas. Además, la distribución de cada variable, condicionada por **cada valor de la otra variable**, debería tener aproximadamente la misma varianza. Sin embargo, como ocurre con la prueba t y el análisis de varianza, los incumplimientos moderados de estos supuestos no son fatales.

A continuación presentamos un ejemplo utilizando el estudio del estrés de los gerentes. Supondremos que los investigadores predijeron una correlación positiva entre la cantidad de empleados supervisados y el estrés, la que será probada a nivel 0,05.

1. Reformule el problema en forma de hipótesis de investigación e hipótesis nula acerca de las poblaciones. Las poblaciones de interés son las siguientes:
Población 1: gerentes como los analizados en este estudio.
Población 2: gerentes para los cuales no existe correlación entre cantidad de empleados su pervisados y estrés.

[2] Dunlap y Myers (1997) encuentran un modo más corto de descubrir la significación de un coeficiente de correlación. Sucede que el r necesario para una significación de nivel 0,05 (dos colas) es muy aproximado a $2/\sqrt{N}$. Por ejemplo, para $N = 5$, necesitaríamos una correlación de 0,89 ($2/\sqrt{5} = 2/2,24 = 0,89$). Dunlop y Myers también nos brindan una manera más corta para lograr una aproximación a la cantidad de participantes necesarios para un poder de entre un 80% y 90%. El tamaño de muestra necesario es simplemente 8 dividido r^2. Por ejemplo, utilizando esta fórmula, para $r = 0,10$, la cantidad de participantes necesaria es $8/0,10^2$, es decir, 800.

La hipótesis nula establece que las dos poblaciones tienen la misma correlación. La hipótesis de investigación establece que la población 1 tiene una correlación mayor que la población 2. (Es decir, la predicción es que la correlación de la población es mayor a 0).

2. Determine las características de la distribución comparativa. Suponiendo que se cumplen los supuestos (en la práctica, con sólo cinco casos sería difícil de determinar), la distribución comparativa es una distribución t con $gl = 3$. (Es decir, $gl = N - 2 = 5 - 2 = 3$).

3. Determine el punto crítico en la distribución comparativa, en el cual la hipótesis nula debería ser rechazada. La tabla t (tabla B-2 del apéndice B) muestra que para una prueba de una cola a nivel 0,05, con 3 grados de libertad, necesitamos una t de al menos 2,353.

4. Determine el valor del estadístico de prueba. Calculamos una correlación de $r = 0{,}88$ y $N = 5$. Aplicando la fórmula para encontrar el t equivalente, obtenemos:

$$t = \frac{(r)\left(\sqrt{N-2}\right)}{\sqrt{1-r^2}} = \frac{(0{,}88)\left(\sqrt{3}\right)}{\sqrt{1-0{,}77}} = \frac{(0{,}88)(1{,}73)}{\sqrt{0{,}23}} = \frac{1{,}52}{0{,}48} = 3{,}17$$

5. Compare los valores obtenidos en los pasos 3 y 4 para decidir si se rechaza la hipótesis nula. El valor t de 3,17, obtenido de la muestra estudiada, es más extremo que el punto t crítico necesario de 2,353. Se rechaza la hipótesis nula y se sostiene la hipótesis de investigación.

Tamaño del efecto y potencia

El coeficiente de correlación es, en sí mismo, una medida del tamaño del efecto. Las reglas de Cohen (1988) para el coeficiente de correlación establecen que el tamaño del efecto de 0,10 es pequeño, de 0,30 es mediano y de 0,50 es grande. La tabla 3-12 nos muestra la potencia aproximada, y la tabla 3-13 el tamaño mínimo de muestra para obtener una potencia del 80% (*véase* también nota al pie 1). Se pueden encontrar tablas más completas en Cohen (1988), pp. 84 95 y 101–102.

Tabla 3-12.
Potencia aproximada de estudios que utilizan el coeficiente de correlación (r) para pruebas de hipótesis con nivel de significación de 0,05.

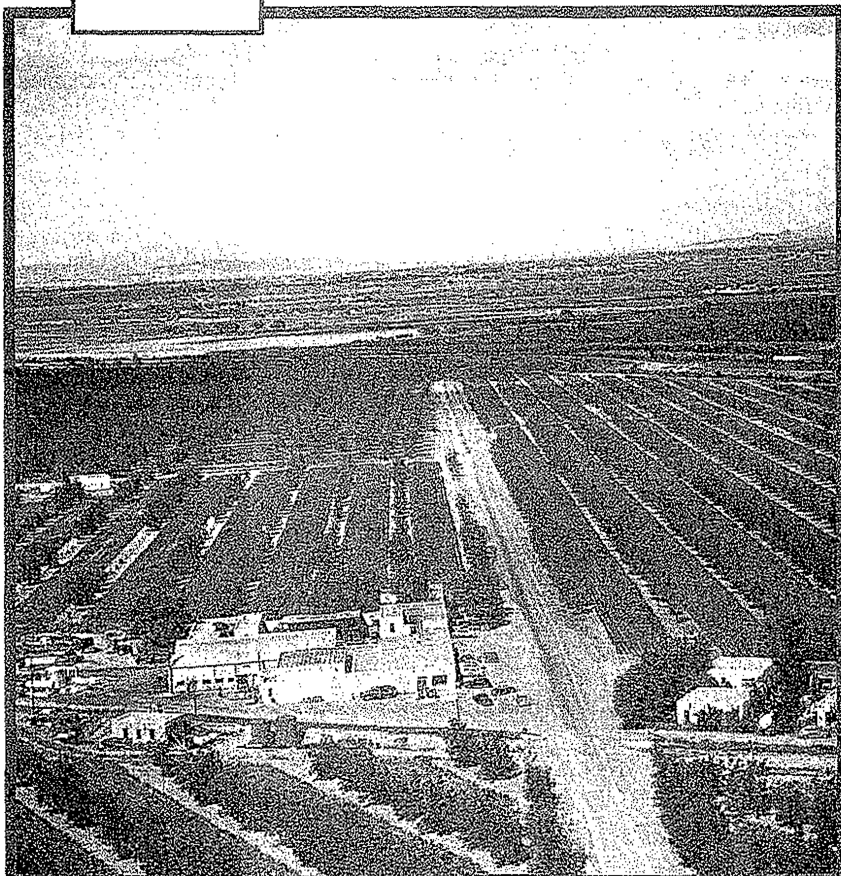
		Tamaño del efecto		
		Pequeño ($r = 0{,}10$)	*Mediano* ($r = 0{,}30$)	*Grande* ($r = 0{,}50$)
Dos colas				
Total N:	10	0,06	0,13	0,33
	20	0,07	0,25	0,64
	30	0,08	0,37	0,83
	40	0,09	0,48	0,92
	50	0,11	0,57	0,97
	100	0,17	0,86	[a]
Una cola				
Total N:	10	0,08	0,22	0,46
	20	0,11	0,37	0,75
	30	0,13	0,50	0,90
	40	0,15	0,60	0,96
	50	0,17	0,69	0,98
	100	0,26	0,92	[a]

[a]Casi 1,00.

Tabla 3-13.
Cantidad aproximada de participantes necesarios para lograr un 80% de potencia en un estudio que utiliza el coeficiente de correlación (r) para probar una hipótesis con nivel de significación de 0,05.

	Tamaño del efecto		
	Pequeño ($r = 0{,}10$)	*Mediano* ($r = 0{,}30$)	*Grande* ($r = 0{,}50$)
Dos colas	783	85	28
Una cola	617	68	22

4 Predicción



Descripción del capítulo

- Terminología relacionada con la predicción bivariada.
- Modelo de predicción bivariada con puntuaciones Z.
- Predicción bivariada con puntuaciones originarias.
- La recta de regresión.
- Error y reducción proporcional del error.
- Otro ejemplo de predicción bivariada.
- Extensión a regresión y correlación múltiples.
- Controversias y limitaciones.
- Los modelos de predicción según se describen en publicaciones científicas.
- Resumen.
- Términos clave.
- Ejercicios.

En este capítulo, y sobre la base de lo aprendido en el capítulo 3, estudiaremos una de las principales aplicaciones prácticas de los métodos estadísticos: realizar predicciones. Normalmente, se recurre a psicólogos de distintas especialidades para solicitar opiniones fundamentadas (y precisas) sobre temas tales como, por ejemplo, cuál es la probabilidad de que el aspirante a un empleo se desempeñe correctamente si se lo contrata, cuánto puede ayudar un programa de lectura a un determinado alumno de tercer grado o cuáles son las probabilidades de que un convicto con posibilidad de salir en libertad condicional cometa un crimen si se lo libera. Aprender los intrincados detalles de la predicción estadística también ayudará a profundizar la comprensión de otros contenidos de la materia, y preparará al alumno para temas fundamentales en cursos de estadística más avanzados.

A lo largo del capítulo analizaremos los procedimientos para realizar predicciones referidas a una variable (como el promedio de calificaciones universitarias), sobre la base de información relacionada con otra variable (como por ejemplo, las calificaciones SAT). Luego veremos cómo estimar la precisión esperada de las predicciones que realizamos utilizando estos procedimientos.

Finalmente, presentaremos situaciones en las que se realizan predicciones referidas a una variable (como el GPA) que se basan en información relacionada con otras dos o más variables (como por ejemplo las calificaciones SAT y el GPA del colegio secundario).

TERMINOLOGÍA RELACIONADA CON LA PREDICCIÓN BIVARIADA

En la **predicción bivariada** (**bivariada** significa "con dos variables"), también llamada regresión bivariada, se utiliza el valor observado de una variable en una persona (por ejemplo, calificaciones SAT) para realizar predicciones sobre el valor de esa persona en otra variable (por ejemplo, GPA universitario). A los fines de recordar la terminología presentada en el capítulo 3, diremos que la variable que ayuda a realizar la predicción (como por ejemplo las calificaciones SAT) se denomina **variable predictora**. (La variable predictora recibe con frecuencia el nombre de **variable independiente**, especialmente si se la considera causa de la otra variable). La variable para la cual se realizan las predicciones (como por ejemplo el GPA universitario) generalmente se denomina **variable dependiente**. (La variable dependiente en una predicción recibe el nombre técnico de **variable criterio**, pero este nombre es poco común en la mayoría de las áreas de investigación psicológica). Usualmente se rotula la variable de predicción con una X y la variable dependiente con una Y. Es decir, se utiliza el valor observado de una persona en X para predecir el valor Y. (La tabla 4-1 resume las distintas denominaciones de las variables).

Ya nos hemos referido a estos dos tipos de variables en nuestra exposición sobre correlación en el capítulo 3. Sin embargo, en ese contexto había relativamente muy poca diferencia con respecto a cuál era cuál, ya que sólo nos interesaba el grado de relación entre ambas. En el contexto de las predicciones, sin embargo, es esencial estar seguro respecto de qué variable se están realizando las predicciones y cuál se está utilizando como ayuda para realizarlas.

MODELO DE PREDICCIÓN BIVARIADA CON PUNTUACIONES Z

Es más simple aprender la predicción bivariada si primero estudiamos la predicción utilizando puntuaciones Z. El **modelo de predicción**, o fórmula, que utilizamos para realizar predicciones con puntuaciones Z es el siguiente: la puntuación Z que se predice para una persona en la variable dependiente se calcula multiplicando un número determinado, denominado **coeficiente de regresión**, por la puntuación Z de esa persona en la variable de predicción.

Dado que estamos trabajando con puntuaciones Z, a las que también se denominan **puntuaciones estándar**, el coeficiente de regresión en este caso recibe el nombre de **coeficiente de regresión estandarizado** y se simboliza con la letra griega "beta" (β). Simbólicamente:

$$\hat{Z}_Y = (\beta)(Z_X) \qquad (4\text{-}1)$$

En esta fórmula, $\hat{Z}_Y$ es el valor predicho de la puntuación Z de una persona en particular, en la variable dependiente Y; el símbolo ô sombrero significa "valor predicho de"; β es el coeficiente de regresión estandarizado; Z_X es la puntuación Z de esa persona en particular en la variable predictora X.

Supongamos que en determinada escuela el coeficiente beta para predecir el GPA universitario (al graduarse) a partir del SAT (al ingresar) es 0,3. Determinada persona que quiere ingresar a ese colegio tiene un SAT que corresponde a dos desvíos estándares por sobre la media (es decir, una

Tabla 4-1.
Denominación de las dos variables en la predicción bivariada.

	Variable a partir de la cual se predice	**Variable que se predice**
Nombre	Variable predictora	Variable dependiente
Nombre alternativo	Variable independiente	Variable criterio
Símbolo	X	Y
Ejemplo	calificaciones SAT	GPA universitario

puntuación Z de +2). La puntuación Z predicha para el GPA de esa persona sería 0,3 por 2, lo que da 0,6. Es decir, la puntuación Z predicha de esa persona para el GPA de su facultad es 0,6 desvíos estándares sobre la media. En símbolos es:

$$\hat{Z}_Y = (\beta)(Z_X) = (0{,}3)(2) = 0{,}6$$

Coeficiente de regresión estandarizado (β) como r

El mejor número para utilizar como beta es el coeficiente de correlación (hecho que puede probarse *matemáticamente utilizando métodos que exceden el alcance de un texto introductorio*). Es decir, en la predicción bivariada, $\beta = r$.

Para comprender mejor el tema, analicemos dos situaciones extremas. Primero, supongamos que no existe correlación alguna entre la variable de predicción y la dependiente. Cuando $r = 0$, conocer el valor de una persona en la variable de predicción no nos ayuda a realizar predicciones; es simplemente irrelevante. Por lo tanto, nuestra mejor predicción es que la persona tendrá un valor en la variable dependiente igual al promedio. Al trabajar con puntuaciones Z, el promedio es siempre 0. Por lo tanto, un coeficiente beta de 0 asegura que cualquiera sea el valor en la variable independiente, la predicción siempre será igual a 0 (ya que 0 veces cualquier número es 0):

$$\text{Cuando } r = 0\text{: } \hat{Z}_Y = (\beta)\,(Z_X) = (0)\,(Z_X) = 0$$

Ahora analicemos la situación extrema en la que existe una correlación perfecta ($r = 1$) entre la variable de predicción y una variable dependiente. Cuando nos encontramos frente a una correlación perfecta, la puntuación Z en la variable predictora es siempre igual a la puntuación Z en la variable dependiente. Cabe recordar lo planteado en el capítulo 3: una correlación significa que los altos coinciden con los altos y los bajos con los bajos. Los valores altos y los bajos son medidos precisamente por las puntuaciones Z, y una correlación perfecta significa que los valores altos coinciden perfectamente con los altos y los bajos perfectamente con los bajos. Cualquier número multiplicado por 1 es igual a sí mismo, y cuando existe una correlación positiva perfecta, beta es 1 (cuando existe una correlación negativa perfecta, beta es = −1):

$$\text{Cuando } r = 1\text{: } \hat{Z}_Y = (\beta)\,(Z_X) = (1)\,(Z_X) = Z_X$$

Por lo tanto, cuando la correlación entre la variable predictora y la dependiente es 0, el mejor número para beta es 0; cuando la correlación es 1, el mejor número para beta es 1. No es sorprendente entonces que en los casos intermedios, cuando r se ubica entre 0 y 1, el mejor número para beta también se ubique entre 0 y 1.

Ejemplo

Analicemos nuevamente el ejemplo del nivel de estrés de los gerentes presentado en el capítulo 3.

En ese ejemplo, la correlación entre la cantidad de empleados supervisados y el nivel de estrés de los gerentes era 0,88; es decir, $r = 0{,}88$. Por lo tanto, B = 0,88, y el modelo para predecir la puntuación Z del nivel de estrés de un gerente es multiplicar 0,88 por la puntuación Z correspondiente a la cantidad de empleados que supervisará el gerente. Supongamos que un nuevo gerente fuera a supervisar a 10 empleados. Esto representaría una puntuación Z de empleados supervisados igual a +1,27. (Cambiamos la puntuación original de 10 a puntuación Z utilizando el procedimiento aprendido en el capítulo 2: $Z = [X - M]/SD$). Así, predeciríamos la puntuación Z del nivel de estrés de este nuevo gerente multiplicando 0,88 por 1,27. El resultado es 1,12, lo que significa que se puede predecir que un gerente que supervisa 10 empleados tendrá un nivel de estrés apenas mayor a 1 desvío estándar por sobre la media. Según la fórmula:

$$\hat{Z}_Y = (\beta)(Z_X) = (0{,}88)(1{,}27) = 1{,}12$$

Por el contrario, supongamos que el nuevo gerente supervisará sólo a 3 empleados. En ese caso, el modelo predeciría una puntuación Z del nivel de estrés igual a 0,88 x (–1,69) (la puntuación Z correspondiente si la cantidad supervisada fuera 3), lo que da un resultado de – 1,49. Es decir:

$$\hat{Z}_Y = (\beta)(Z_X) = (0{,}88)(-1{,}69) = -1{,}49$$

¿Por qué a veces la predicción se denomina regresión?

Los psicólogos usualmente se refieren a este tipo de predicción como **regresión**. El término proviene del hecho de que cuando existe una correlación menos que perfecta entre dos variables, la puntuación Z de la variable dependiente es una fracción de la puntuación Z de la variable predictora (la fracción equivale al valor de *r*). Como resultado, la puntuación Z de la variable dependiente se encuentra más cerca de su media. Es decir, sufre una regresión o retorno hacia un Z de 0.

En el ejemplo del estrés sufrido por los gerentes, el nuevo gerente que iba a supervisar a 10 empleados tiene una puntuación Z correspondiente a empleados supervisados igual a 1,27, pero la puntuación Z predicha para el nivel de estrés ha "sufrido una regresión" a sólo 1,12.

PREDICCIÓN BIVARIADA CON PUNTUACIONES ORIGINALES

En esta sección analizamos dos formas de realizar predicciones utilizando puntuaciones originales.

Convertir puntuaciones originales en Z. Realizar las predicciones. Convertir puntuaciones Z en originales

Una manera de realizar predicciones con puntuaciones originales es la siguiente:

1. Convertir la puntuación original de la variable de predicción (*X*) en una puntuación Z (Z_X).
2. Multiplicar beta (el coeficiente de correlación) por esa puntuación Z (Z_X) para obtener la puntuación Z predicha de la variable dependiente ($\hat{Z}_Y$).
3. Convertir la puntuación Z predicha de la variable dependiente ($\hat{Z}_Y$) en una puntuación original ($\hat{Y}$).

En el ejemplo del estrés sufrido por los gerentes, cuando queríamos predecir el nivel de estrés de un gerente que supervisaba 10 empleados, primero convertíamos 10 en puntuación Z (Z_X = 1,27) (paso 1). Luego encontrábamos la puntuación Z predicha para el nivel de estrés multiplicando beta por el punto Z correspondiente a la cantidad de empleados supervisados. (0,88 y 1,27 daba una puntuación Z predicha, $\hat{Z}_Y$, de 1,12) (paso 2). El paso 3 (que no realizamos anteriormente) es convertir esa puntuación Z predicha de 1,12 nuevamente en puntuación original. Utilizando la fórmula del capítulo 2 para convertir una puntuación Z en una puntuación original, el resultado es 8,92 ($\hat{Y}$ = 8,92). Es decir, utilizando el procedimiento de regresión, predecimos que un gerente que supervisa 10 empleados tendrá un nivel de estrés de 8,92.

La tabla 4-2 describe estos pasos aplicados al otro ejemplo (el ejemplo del gerente que supervisaría a 3 personas).

Tabla 4-2.
Resumen de los pasos que se deben seguir para realizar predicciones con puntuaciones originales, a través de las conversiones de original a Z y de Z a original, utilizando fórmulas y tomando un ejemplo como base.

Paso	Fórmula	Ejemplo
1	$Z_X = (X - M_X)/SD_X$	$Z_X = (3 - 7)/2{,}37 = -1{,}69$
2	$\hat{Z}_Y = (\beta)(Z_X)$	$\hat{Z}_Y = (0{,}88)(-1{,}69) = -1{,}49$
3	$\hat{Y} = (SD_Y)(\hat{Z}_Y) + M_Y$	$\hat{Y} = (2{,}61)(-1{,}49) + 6 = 2{,}11$

Al realizar los tres pasos indicados anteriormente, se debe poner especial atención en utilizar la media y el desvío estándar de la variable correspondiente al pasar de puntuaciones originales a puntuaciones Z y de puntuaciones Z a puntuaciones originales. En el paso 1, se trabaja sólo con el valor, la media y el desvío estándar de la variable predictora (X). En el paso 3, se trabaja sólo con el valor, la media y el desvío estándar de la variable dependiente (Y).

Predicción directa de puntuación original a puntuación original

Un procedimiento alternativo reduce a una sola fórmula el proceso de los tres pasos anteriormente utilizado. Esta sola fórmula toma en cuenta automáticamente la conversión en puntuaciones Z y de puntuaciones Z (pasos 1 y 3). Es decir, si en el modelo de predicción se incluyen las fórmulas para la conversión en puntuaciones Z, y de puntuaciones Z, y se realizan algunas manipulaciones algebraicas, se puede lograr una sola **fórmula de predicción con puntuaciones originales**:

$$\hat{Y} = a + (b)(X) \tag{4-2}$$

Esta fórmula hace hincapié en dos términos que aún no hemos analizado, b y a. b es el **coeficiente de regresión para puntuaciones originales**, es similar a β, el coeficiente de regresión estandarizado, excepto que b se utiliza sólo con puntuaciones originales y no es igual al coeficiente de correlación. a es la **constante de regresión**, se agrega al valor predicho en la variable dependiente de puntuaciones originales, para tomar en cuenta las medias de las distribuciones de puntuaciones originales. (Trabajando con puntuaciones Z, no es necesario utilizar la constante de regresión debido a que las medias de las puntuaciones Z de las variables siempre son iguales a 0).

El coeficiente de regresión para puntuaciones originales (b) y la constante de regresión (a) pueden calcularse directamente conociendo las medias y los desvíos estándares de las dos variables, y beta (que en el caso de las predicciones bivariadas es r):

$$b = (\beta)\left(\frac{SD_Y}{SD_X}\right) \tag{4-3}$$

$$a = M_Y - (b)(M_X) \tag{4-4}$$

Según nuestro ejemplo del estrés sufrido por los gerentes, $r = M_X = 7$, $SD_X = 2{,}37$, $M_Y = 6$, y $SD_Y = 2{,}61$. Entonces,

$$b = (\beta)\left(\frac{SD_Y}{SD_X}\right) = (0{,}88)\left(\frac{2{,}61}{2{,}37}\right) = (0{,}88)(1{,}10) = 0{,}97$$

$$a = M_Y - (b)(M_X) = 6 - (0{,}97)(7) = 6 - 6{,}79 = -0{,}79$$

$$Y = a + (b)(X) = -0{,}79 + (0{,}97)(X)$$

Si un gerente supervisa a 10 individuos, el nivel de estrés predicho será igual a:

$$\hat{Y} = -0{,}79 + (0{,}97)(X) = -0{,}79 + (0{,}97)(10) = -0{,}79 + 9{,}7 = 8{,}91$$

Si supervisa a 3 personas:

$$\hat{Y} = -0{,}79 + (0{,}97)(X) = -0{,}79 + (0{,}97)(3) = -0{,}79 + 2{,}91 = 2{,}12$$

(Como puede observarse, teniendo en cuenta los redondeos, los resultados coinciden con los cálculos realizados utilizando el método de tres pasos, que implica conversión de original a Z, predicción, conversión de Z a original).

De un modo más general, analicemos el significado de b y a según lo ilustra el siguiente ejemplo: el coeficiente de regresión para puntuaciones originales (b) de 0,97 significa que cada aumento de una persona supervisada está ligado a un aumento de 0,97 puntos sobre el valor que se predice para el nivel de estrés de los gerentes. Si se supervisan dos personas, se multiplica 0,97 por 2; si son tres, 0,97 por 3.

La constante de regresión (a) de –0,79 significa que, además, se ajusta la predicción restando *0,79 puntos a la escala de estrés, cualquiera sea la cantidad de empleados.* Justamente se trata de una constante porque siempre se utiliza el mismo valor.

La constante de regresión de –0,79 también indica que si X es 0, el registro de estrés será de –0,79. (Sin embargo, en este caso X es la cantidad de empleados supervisados, y resulta improbable que un gerente no supervise a ningún empleado, lo cual es un hecho afortunado, ya que también es improbable que alguien pudiera tener menos de 0 estrés).

LA RECTA DE REGRESIÓN

Un modelo de predicción puede visualizarse como una recta en un gráfico, en el que el eje horizontal representa los valores de la variable predictora y el eje vertical representa los valores predichos de la variable dependiente. (El gráfico se dibuja del mismo modo que los diagramas de dispersión aprendidos en el capítulo 3). La recta a la que nos referíamos se llama **recta de regresión,** y representa la relación entre los valores de la variable predictora y los valores predichos en la variable dependiente. La figura 4-1 grafica la recta de regresión correspondiente al ejemplo de los empleados supervisados (variable predictora) y el nivel de estrés de los gerentes (variable dependiente). Siguiendo la recta de regresión se puede encontrar el nivel de estrés predicho a partir de cualquiera de las cantidades de empleados supervisados. Las líneas punteadas indican las predicciones calculadas para los gerentes que supervisaban 3 y 10 personas.

Pendiente de la recta de regresión

Es particularmente interesante la inclinación de la recta de regresión, a la cual se denomina **pendiente.** La pendiente indica cuánto se eleva la recta por cada unidad de incremento de la variable predictora. En el ejemplo de la figura 4-1, la línea se eleva 0,97 puntos de estrés por cada perso-

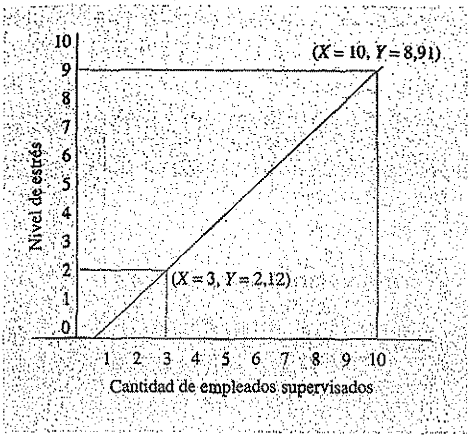


Figura 4-1. Recta de regresión correspondiente al ejemplo en el que se utilizaron puntuaciones originales para predecir el estrés de los gerentes, con la indicación de los niveles predichos de estrés para gerentes que supervisan 3 y 10 empleados.

na adicional supervisada. De hecho, la pendiente de la línea es exactamente *b*, el coeficiente de regresión.

Esta equivalencia entre la pendiente de la recta de regresión y *b* acentúa el hecho de que un coeficiente de regresión sirve como una especie de **razón de cambio** entre la variable predictora y la dependiente. Es decir, el coeficiente de regresión indica cuántas unidades predichas de la variable dependiente se obtienen por una cantidad dada de unidades de la variable de predicción. Es como saber que en determinado día, con un dólar canadiense se adquieren cinco francos franceses. (De todos modos, no se debe abusar de esta analogía. Al cambiar dinero, realizamos una transacción más o menos exacta. Con los modelos de predicción, el cambio es entre una cifra real en la variable predictora y una cantidad predicha en la variable dependiente. Excepto en el caso de una correlación perfecta, la predicción no será exacta).

Cómo trazar la recta de regresión

El primer paso es establecer los ejes y los rótulos del gráfico, conforme a lo aprendido en el capítulo 3, para construir un diagrama de dispersión. La recta de regresión es una línea recta, de modo que sólo se necesita calcular la ubicación de cualquier par de puntos y dibujar la recta que pasa a través de ellos. Los pasos que se deben seguir son cuatro, y los ilustraremos con el ejemplo del nivel de estrés de los gerentes:

1. Dibujar y rotular los ejes para un diagrama de dispersión de dos variables, según se describe en el capítulo 3, con la variable predictora en el eje horizontal. (Se podría utilizar la regla mnemotécnica "lo conocido forma una **base** estable para lo que se predice o prevé en las **alturas**"). La figura 4-2a ilustra este paso aplicado al ejemplo del nivel de estrés de los gerentes.

2. Escoger cualquier valor de la variable predictora, calcular el correspondiente valor predicho en la variable dependiente y marcar el punto en el gráfico. Si se selecciona un valor de la variable predictora igual a 2, el valor predicho en la variable dependiente, según nuestro ejemplo, será: 0,79+(0,97x 2) = 1,15. En la figura 4-2b se ha marcado este punto ($X = 2, \hat{Y} = 1,15$).

3. Repetir el paso 2 comenzando con cualquier otro valor de la variable predictora. (Se podrá dibujar la recta de forma más precisa si se escoge un valor de la variable predictora bastan-

te diferente del primero). Según el ejemplo que estamos utilizando, si se selecciona un valor de la variable predictora igual a 8, la puntuación Y predicha de la variable dependiente será $-0{,}79 + (0{,}97 \times 8) = 6{,}97$. En la figura 4-2c se ha marcado este punto ($X = 8$, $\hat{Y} = 6{,}97$).

(Cabe recordar entonces que, para trazar una recta de regresión, se seleccionan arbitrariamente dos valores cualesquiera de X y se calcula el valor predicho de Y correspondiente a cada uno de ellos; por lo tanto, se puede trazar una recta de regresión sin necesidad de contar con ningún valor de X en particular).

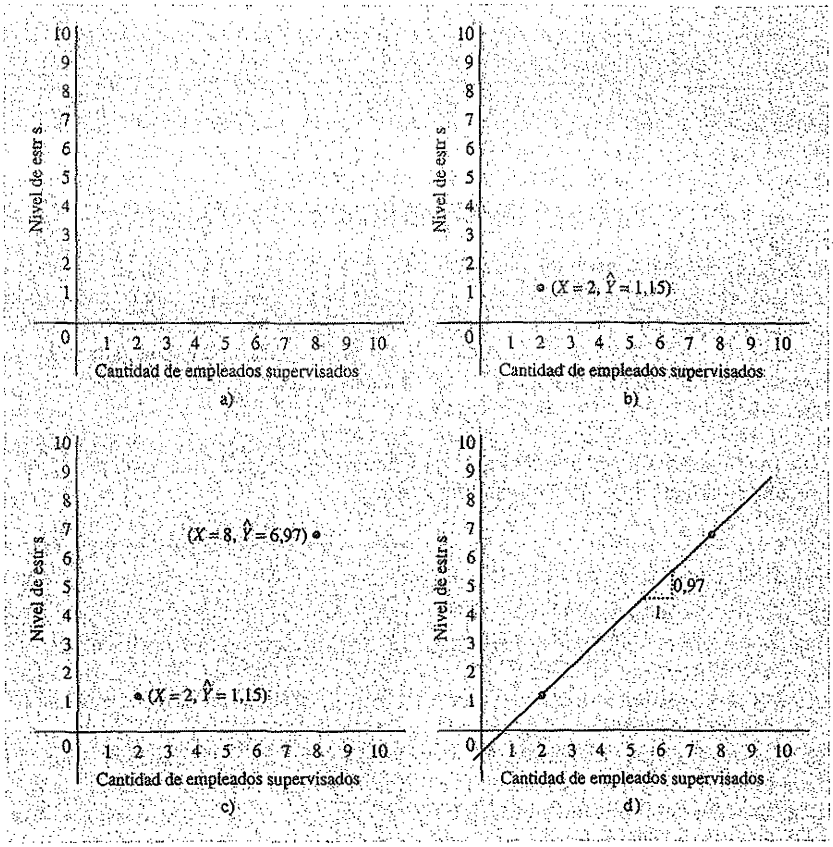


Figura 4-2. Pasos que se deben seguir para trazar una recta de regresión utilizando el ejemplo del nivel de estrés de los gerentes. a) Se dibujan y rotulan los ejes; b) se marca el punto cuyas coordenadas son un valor de la variable predictora (2) y el correspondiente valor predicho calculado para la variable dependiente (1,15); c) se marca un punto cuyas coordenadas son otro valor de la variable de predicción (8) y su correspondiente valor predicho calculado para la variable dependiente (6,97), y d) se dibuja una recta que pase por los dos puntos marcados. El gráfico también indica que por cada unidad de incremento de X, la recta se eleva 0,97 unidades.

4. Dibujar la recta que pasa por los dos puntos marcados. La figura 4-2d muestra la recta.

Se puede controlar la precisión de la línea trazada calculando cualquier otro tercer punto. Un punto fácil de localizar es el punto donde $X = 0$. Cuando $X = 0$, el valor predicho de Y es la constante de regresión (a). (Cuando $X = 0$, $(b)(X) = 0$; por lo tanto, lo único que queda de la fórmula de regresión es a). Frecuentemente, el diagrama de dispersión se realiza de forma tal que el eje vertical esté ubicado donde $X = 0$. En ese caso, el punto en el que la línea de regresión corta el eje vertical es el punto donde el valor predicho en Y es igual a a. Por esta razón, la constante de regresión a veces también se denomina **ordenada al origen** (la ordenada del punto donde la recta de regresión intersecta o corta al eje Y).

Para mayor control, en cuanto a la precisión de la recta trazada, es posible verificar si la pendiente coincide con b, es decir, cuánto se eleva la recta por cada unidad de incremento de la variable predictora. La figura 4-2d muestra con líneas punteadas que la pendiente es 0,97: por cada unidad de incremento de X, la recta se eleva 0,97 unidades.

ERROR Y REDUCCIÓN PROPORCIONAL DEL ERROR

¿Cuán precisas son las predicciones que se realizan utilizando los procedimientos que hemos descripto? Normalmente uno predice el futuro y no existe modo de saber con seguridad qué es lo que sucederá. Pero, sin embargo, sí se puede realizar una estimación.

La estimación puede realizarse analizando cuán preciso hubiera sido el modelo de predicción de habérselo utilizado para realizar "predicciones" de los valores con los que se calculó el coeficiente de correlación en primer lugar. Es decir, primero se crea una norma de predicción calculando el coeficiente de correlación con los valores observados de un grupo de individuos en particular que hayan sido estudiados. Luego se utiliza esta norma de predicción para realizar "predicciones" para esos mismos individuos. Se realizan "predicciones" para cada individuo, incluyendo el valor X de esa persona dentro de la norma de predicción y calculando el valor Y "predicho" para esa persona. (Hemos puesto las palabras "predicción" y "predicho" entre comillas porque en realidad ya se conoce el valor Y de cada persona. Se está utilizando la norma de predicción pero para predecir algo que, en principio, ya se conoce). Una vez obtenido el valor Y "predicho" para cada individuo del grupo que originalmente se analizó, se pueden comparar esos valores Y "predichos" con los valores Y observados de los individuos analizados. Si la norma de predicción es buena, entonces los valores Y "predichos" deberían ser muy similares a los valores Y observados.

Por ejemplo, en el caso de los gerentes, no hay forma de saber a ciencia cierta cuán precisas serán las predicciones del nivel de estrés de nuevos gerentes. Pero uno puede preguntarse cuán precisas hubieran sido esas predicciones si se hubiera utilizado este modelo para predecir el nivel de estrés de los gerentes que ya se han estudiado. Analicemos los cinco gerentes utilizados como ejemplo en el capítulo 3. La correlación entre cantidad de empleados supervisados y nivel de estrés calculado en ese caso era de 0,88, y siguiendo los pasos para convertir esta información en una norma de predicción con puntuaciones originales, descubrimos que $\hat{Y} = -0,79 + (0,97)(X)$.

Una vez realizado lo anterior, ahora se puede aplicar esa norma nuevamente a estos mismos cinco gerentes; por ejemplo, el primero de los cinco gerentes supervisaba a 6 personas y sufría un nivel de estrés de 7. Aplicando la norma de predicción, el nivel de estrés "predicho" para este gerente sería de $-0,79 + (0,97)(6)$, lo que da como resultado un nivel de estrés "predicho" de 5,03.

Las primeras tres columnas de la tabla 4-3 indican la cantidad de empleados supervisados, los niveles reales de estrés y los valores de estrés "predichos" utilizando el modelo de predicción. Se puede observar que los valores de estrés "predichos" son moderadamente cercanos a los valores observados de estrés. (Ahora que hemos explicado el tema, dejaremos de poner comillas a las palabras "predicción" y "predicho" al referirnos a la utilización del modelo de predicción para calcular Y con respecto a cada persona del grupo original de individuos analizados. Tampoco hemos utilizado comillas en la tabla 4-3, pero cabe recordar que éstas predicciones se están realizando con individuos cuyos *valores Y observados ya conocemos*. El *propósito* de desarrollar este proceso no es conocer más acerca de los individuos originalmente analizados, sino más bien controlar la precisión de nuestra norma de predicción).

Error y error cuadrático

El siguiente paso es utilizar las predicciones referidas a las personas originalmente analizadas para determinar la precisión de la norma de predicción. Para ello, primero se calcula cuán alejadas están las predicciones realizadas, utilizando la norma de predicción, de los valores observados de los individuos analizados originalmente. Esta distancia es lo que denominamos **error**, es decir, para cada individuo, el error es el valor observado menos el valor predicho.

El procedimiento siguiente es elevar cada error al cuadrado, obteniendo **errores cuadráticos**. Es decir:

$$Error^2 = \left(Y - \hat{Y}\right)^2 \qquad (4\text{-}5)$$

Utilizar errores cuadráticos soluciona el problema de que algunos errores sean números positivos (la predicción fue menos que la observación) y otros sean números negativos (la predicción superó a la observación). Si no se elevan los errores al cuadrado, cuando finalmente se sumen, los errores positivos y negativos se cancelarán entre sí. (La misma situación se planteó en el capítulo 2 cuando trabajamos con desvíos con respecto a la media).

Los errores y errores cuadráticos correspondientes a los gerentes del ejemplo aparecen en las últimas dos columnas de la tabla 4-3

Interpretación gráfica del error

La figura 4-3 muestra el diagrama de dispersión que representa el ejemplo del nivel de estrés sufrido por los gerentes, con su correspondiente recta de regresión. En el gráfico, los cinco pares de valores observados están indicados con puntos; los valores de estrés predichos para todos los va-

Tabla 4-3.
Cálculo del error y del error cuadrático utilizando puntuaciones originales del ejemplo del nivel de estrés sufrido por los gerentes (datos ficticios).

Empleados Supervisados	Nivel de Estrés		Error	Error²
	Observado	*Predicho*		
X	Y	$\hat{Y}$	$Y-\hat{Y}$	$(Y-\hat{Y})^2$
6	7	5,03	1,97	3,88
8	8	6,97	1,03	1,06
3	1	2,12	−1,12	1,25
10	8	8,91	−0,91	0,83
8	6	6,97	−0,97	0,94
				Suma = 7,96

lores correspondientes a la cantidad de empleados supervisados se encuentran a lo largo de la recta de regresión. Por lo tanto, el error correspondiente a cualquier gerente en particular está dado por la distancia vertical entre el punto correspondiente al valor observado de ese gerente y la recta de regresión. Se han dibujado líneas de puntos para indicar el error en cada caso.

Reducción proporcional del error

Ahora bien, ¿cuál es la utilidad de los errores cuadráticos? El modo más común de considerar la precisión de nuestro modelo de predicción es comparar la cantidad de error cuadrático, que probablemente existiría utilizando el modelo de predicción, con la cantidad de error cuadrático que existiría sin utilizar el modelo. La estrategia a seguir es la siguiente: a) calcular la cantidad de error cuadrático que existiría si predijéramos utilizando el modelo de predicción; b) calcular la cantidad de error cuadrático que existiría prediciendo sin utilizar el modelo, y c) comparar las dos cantidades cuadráticas.

La cantidad de error cuadrático que existiría utilizando el modelo de predicción es la **suma de los errores cuadráticos**. Es decir, sólo hay que sumar los errores cuadráticos de todos los individuos originalmente analizados. En el ejemplo del nivel de estrés de los gerentes, equivaldría a la suma de la última columna de la tabla 4-3, que resulta ser 7,96. La suma de los errores cuadráticos se abrevia SS_{Error}. Por lo tanto, en nuestro ejemplo, $SS_{Error} = 7{,}96$.

¿Cómo se calcula la cantidad de error cuadrático sin el modelo? Supongamos que no se pudiera utilizar el modelo de predicción, es decir, que no se pudiera tener en cuenta el nivel de cada individuo en la variable de predicción. La regla general es la siguiente: si no es posible utilizar el modelo de predicción, la predicción más acertada será predecir con la media el valor de cada individuo.

Tomando nuevamente el ejemplo de los gerentes, si no se supiera nada sobre la variable de predicción (cantidad de empleados que supervisa cada gerente), la mejor estrategia sería predecir que un gerente tendrá un nivel de estrés igual a la media. Es decir, si no se pudiera utilizar información sobre cantidades de empleados supervisados, la mejor estrategia sería predecir un valor de estrés de 6 para cada gerente.

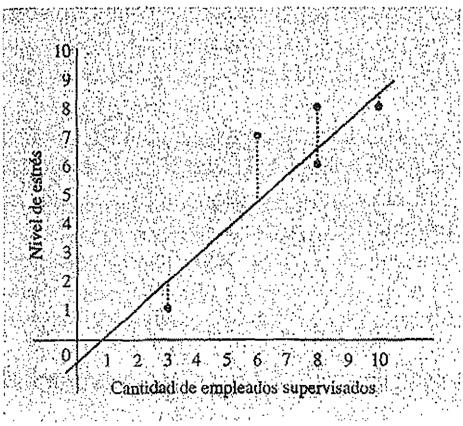


Figura 4-3. Diagrama de dispersión para el ejemplo del nivel de estrés de los gerentes, con la recta de regresión trazada y líneas punteadas que indican los errores (distancia vertical desde el valor observado, indicado por un punto, y el valor predicho, indicado por el correspondiente punto ubicado sobre la recta de regresión).

Analicemos otro ejemplo. Supongamos que se intentara predecir el GPA universitario de una persona, pero que no fuera posible utilizar un modelo basado en el SAT o cualquier otra variable de predicción. En ese caso, la mejor apuesta sería predecir que el GPA universitario de esa persona sería el GPA universitario medio de los alumnos de esa facultad.

Por lo tanto, **la cantidad de error cuadrático al predecir sin un modelo es la cantidad de error cuadrático calculado al predecir con la media cada valor observado**. Es importante recordar que el error, en general, es el valor observado menos el valor predicho. Cuando el valor predicho es la media, el error es el valor observado menos la media, el error cuadrático es el cuadrado de ese número, y la suma de estos errores cuadráticos es el **error cuadrático total al predecir con la media**; llamamos a este número SS_{Total}.

(Lo que ahora llamamos SS_{Total} es lo mismo que llamamos SS en el capítulo 2, como parte del cálculo de la varianza. Definimos SS como la suma de los desvíos cuadráticos con respecto a la media. Un desvío con respecto a la media es igual al valor observado menos la media, que es exactamente lo mismo que el error que resulta cuando la predicción es la media).

Ahora conocemos a ambos, (a) la suma de los errores cuadráticos al predecir utilizando el modelo de predicción (SS_{Error}) y (b) la suma de los errores cuadráticos al predecir utilizando la media (SS_{Total}). La ventaja del modelo de predicción es la ventaja de SS_{Error} con respecto a SS_{Total}, es decir, la medida en la cual cometemos menos errores utilizando el modelo de predicción que utilizando la media. Con un buen modelo de predicción, SS_{Error} debería ser menor que SS_{Total}.

La comparación mencionada en el párrafo anterior es un indicador de la precisión del modelo de predicción, y se denomina **reducción proporcional del error**. Para calcular la reducción proporcional del error, primero se debe encontrar la reducción del error, es decir, la diferencia entre el error cuadrático al predecir utilizando la media (SS_{Total}) y el error cuadrático utilizando el modelo de predicción (SS_{Error}). Es decir, se calcula $SS_{\text{Total}} - SS_{\text{Error}}$. Luego, el resultado, que es la reducción del error cuadrático que se logra utilizando el modelo de predicción, se divide por la cantidad total del error. Se representa mediante la fórmula:

$$\text{Reducción proporcional del error} = \frac{SS_{Total} - SS_{Error}}{SS_{Total}} \tag{4-6}$$

Es decir, utilizar la media para predecir no es un método muy preciso porque produce mucho error. Mediante el cálculo descripto se comprueba cuánto mejor se puede realizar la predicción. La proporción de error cuadrático en el que se incurriría utilizando la media se reduce utilizando la norma de predicción.

Analicemos una situación en la que el modelo de predicción no produce ninguna mejora en comparación con la predicción que se realiza utilizando la media. En este caso, SS_{Error} es igual a SS_{Total} (SS_{Error} nunca puede ser menor que SS_{Total}). El modelo de predicción no ha reducido el error ($SS_{\text{Total}} - SS_{Error} = 0$) y ha reducido un 0% el error total ($0/SS_{\text{Total}} = 0$).

Ahora analicemos una situación en la que el modelo de predicción realiza predicciones perfectas, sin ningún error. El modelo de predicción ha reducido el error en un 100%. (Expresado con la ecuación, si $SS_{\text{Error}} = 0$, entonces el numerador será $SS_{\text{Total}} - 0$, ó SS_{Total}; dividir SS_{Total} por SS_{Total} da 1, o sea un 100%).

En la mayoría de los casos reales, la reducción proporcional del error se encuentra entre el 0% y el 100%.

Ejemplo

La tabla 4-4 indica las predicciones con puntuaciones originales, errores, errores cuadráticos, sumas de errores cuadráticos y reducciones proporcionales del error en el estudio del nivel de estrés de los gerentes. De la tabla se desprende que a través del modelo de predicción se reduce en un 77% el error en el que se incurriría utilizando la media como predictor.

Reducción proporcional del error como r^2

La reducción proporcional del error siempre es igual al cuadrado del coeficiente de correlación. Es decir:

$$\text{Reducción proporcional del error} = r^2 \tag{4-7}$$

Debido a esta equivalencia, r^2 se utiliza generalmente como símbolo de la reducción proporcional del error.

Por ejemplo, en el estudio del nivel de estrés de los gerentes, el coeficiente de correlación era 0,88, y 0,88 al cuadrado es 0,77. Es decir, $r^2 = 0{,}77$. Este número (0,77) es exactamente igual al que acabamos de calcular a través de los valores predichos, errores, errores cuadráticos, sumas de errores cuadráticos y reducción proporcional del error cuadrático.

Calculamos la reducción proporcional del error tan laboriosamente sólo para facilitar la comprensión de este importante concepto. (Para incorporar la lógica, recomendamos realizar el mismo procedimiento con algunos ejemplos adicionales, como los que aparecen en los ejercicios). Sin embargo, en un caso real de investigación, se utilizaría el procedimiento simple de elevar el coeficiente de correlación al cuadrado.

La reducción proporcional del error a veces se denomina **proporción de varianza explicada**. Se utiliza este nombre porque SS_{Total} es una especie de medida de varianza a partir de la media de la variable dependiente, y está muy relacionada con la varianza de la variable dependiente. (SS_{Total} es lo mismo que SS en la fórmula de varianza: es el número que al ser dividido por N da como resultado la varianza). La reducción proporcional del error indica cuánto disminuye la SS_{Total} o cuánto explica el modelo de predicción de esa SS_{Total}. Por lo tanto, la reducción proporcional del error es también la proporción en que se reduce ese cierto tipo de varianza.

Tabla 4-4.
Cálculo de la reducción proporcional del error en el ejemplo del nivel de estrés de los gerentes (datos ficticios).

Observación	**Predicción utilizando la media**			**Utilizando el modelo de predicción**		
Y	*Media*	*Error*	*Error*^{2}	$\hat{Y}$	*Error*	*Error*^{2}
7	6	1	1	5,03	1,97	3,88
8	6	2	4	6,97	1,03	1,06
1	6	–5	25	2,12	–1,12	1,25
8	6	2	4	8,91	– 0,91	0,83
6	6	0	0	6,97	– 0,97	0,94
		$SS_{\text{Total}} = 34$			$SS_{\text{Error}} = 7{,}96$	

$$\text{Reducción proporcional del error} = \frac{SS_{\text{Total}} - SS_{\text{Error}}}{SS_{\text{Total}}} = \frac{34 - 7{,}96}{34} = \frac{26{,}04}{34} = 0{,}77$$

Interpretación gráfica de la reducción proporcional del error

Supongamos que se predijera la media para cada valor. En un gráfico, la línea que represente estas predicciones (todas referidas a la media) sería una recta horizontal. No importa cuál sea el valor de la variable de predicción, la predicción en la variable dependiente es la misma, la media.

La figura 4-4 representa el diagrama de dispersión del ejemplo del nivel de estrés de los gerentes. También ilustra la recta de regresión calculada con el modelo de predicción y la recta horizontal de la predicción utilizando la media. Se puede observar que, en la mayoría de los casos, la recta de regresión está más cerca del punto que la recta horizontal. Es decir, la recta basada en el modelo de predicción, generalmente está más cerca de los puntos que la recta basada en la predicción por la media. La reducción proporcional del error puede considerarse como la medida en la que la precisión de la recta de regresión es mayor que la precisión de la recta horizontal[1].

OTRO EJEMPLO DE PREDICCIÓN BIVARIADA

Apliquemos ahora los distintos aspectos de la predicción bivariada al ejemplo del experimento ficticio del capítulo 3, que indaga sobre la memoria. La tabla 4-5 indica las medias y los desvíos estándares de las dos variables, así como también la correlación entre ellas.

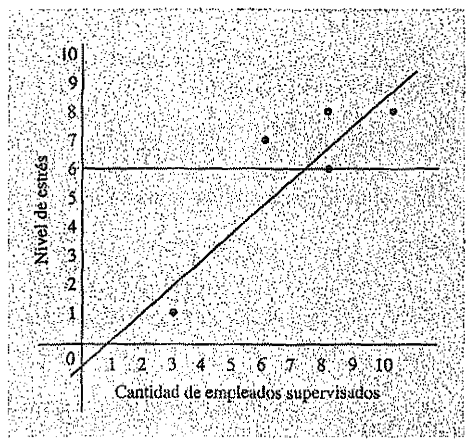


Figura 4-4. Diagrama de dispersión correspondiente al ejemplo del nivel de estrés de los gerentes, que muestra la recta de regresión utilizando el modelo de predicción y la recta horizontal que representa las predicciones realizadas utilizando la media. Los puntos que representan los valores observados, en general, se encuentran más cerca de la recta de regresión que de la recta horizontal.

[1] Existe otra forma menos común de medir la precisión del modelo de predicción, denominada **error estándar de estimación.** Indica, aproximadamente, la distancia promedio entre los puntos y la línea de regresión. Expresado con palabras, es la raíz cuadrada del promedio de los errores cuadráticos, en símbolos $\sqrt{(SS_{Error}/N)}$. (El error estándar de estimación, como indicador de la variación de los valores con respecto a lo que se esperaría aplicando la norma de predicción, es un método paralelo a utilizar el desvío estándar como indicador del desvío típico de los valores con respecto a la media). Sin embargo, el error estándar de estimación rara vez se menciona en los artículos de investigación psicológica. Por lo tanto, en este libro no nos concentramos en ese concepto. No se debe confundir el error estándar de estimación con lo que a menudo se denomina simplemente "error estándar" (técnicamente este *último* es el "error estándar del coeficiente de correlación" o el "error estándar del coeficiente de regresión"), que está relacionado con la significación estadística, y que trataremos en el capítulo 7.

Al utilizar puntuaciones Z, el modelo de predicción consistirá en multiplicar beta, que es 0,68 (igual a r), por la puntuación Z correspondiente a la cantidad de exposiciones.

$$\hat{Z}_Y = (\beta)(\hat{Z}_X) = (0{,}68)(\hat{Z}_X)$$

Supongamos que una persona sea expuesta siete veces a cada palabra. Siete es igual a una puntuación Z de 1,09. Entonces se podría predecir que la puntuación Z de esa persona en la variable "palabras recordadas" sería de 0,68 por 1,09. El resultado es una puntuación Z predicha de 0,74 para las palabras recordadas. Supongamos que otra persona observará cada palabra sólo cuatro veces (una puntuación Z de –0,22 para las 4 exposiciones). En este caso, se predeciría una puntuación Z de –0,15 para las palabras recordadas. Es decir, 0,68 x –0,22 = –0,15. Los dos ejemplos se representan por medio de las siguientes fórmulas:

Para $Z_X = 1{,}09$: $\hat{Z}_Y = (\beta)\,(Z_X) = (0{,}68)\,(1{,}09) = 0{,}74$
Para $Z_X = -0{,}22$: $\hat{Z}_Y = (\beta)\,(Z_X) = (0{,}68)\,(-0{,}22) = 0{,}15$

Sin embargo, cabe recordar que existen dos métodos. Primero, se puede proceder como acabamos de hacerlo. Se puede convertir la puntuación original de la variable de predicción en puntuación Z, realizar la predicción y luego convertir la puntuación Z predicha de la variable dependiente en una puntuación original. En cuanto al resultado del primer ejemplo, una puntuación Z predicha de 0,74 para las palabras recordadas es equivalente a una puntuación original de 7,2 palabras recordadas. (La media de 5,6 más el producto de Z de 0,74 por el desvío estándar de 2,1). Similarmente, una puntuación Z predicha de –0,15 es equivalente a una puntuación original de 5,3 palabras. Es decir, 5,6 + (2,1 x [–0,15]) = 5,3.

Otra alternativa sería utilizar el modelo de predicción con puntuaciones originales para ahorrar algunos pasos. En ese caso, el cálculo sería el siguiente:

$$b = (\beta)\left(\frac{SD_Y}{SD_X}\right) = (0{,}68)\left(\frac{2{,}1}{2{,}29}\right) = (0{,}68)(0{,}92) = 0{,}63$$

$$a = M_Y - (b)(M_X) = 5{,}6 - (0{,}63)(4{,}5) = 5{,}6 - 2{,}84 = 2{,}76$$

$$\hat{Y} = a + (b)(X) = 2{,}76 + (0{,}63)(X)$$

Si una persona accede a siete exposiciones:

$$\hat{Y} = 2{,}76 + (0{,}63)(X) = 2{,}76 + (0{,}63)(7) = 2{,}76 + 4{,}41 = 7{,}17$$

Tabla 4-5.
Medias y desvíos estándares del experimento acerca del efecto de la cantidad de exposiciones sobre la cantidad de palabras recordadas (datos ficticios).

	Cantidad de exposiciones (variable predictora)	**Cantidad recordada (variable dependiente)**
Mean	4,5	5,6
Standard deviation	2,29	2,1
Correlation	$r = 0{,}68$	

Si una persona accede a cuatro exposiciones:

$$\hat{Y} = 2{,}76 + (0{,}63)(X) = 2{,}76 + (0{,}63)(4) = 2{,}76 + 2{,}52 = 5{,}28$$

(Los resultados concuerdan con las cifras más redondeadas que calculamos utilizando el método de transformación de puntuaciones originales a Z, predicción, y transformación de Z a puntuaciones originales).

La figura 4-5 es un gráfico que representa las dos variables y la recta de regresión correspondiente a la fórmula de predicción, junto con las líneas punteadas, que indican las dos predicciones aquí calculadas.

¿Qué podemos decir sobre la precisión de la predicción? La tabla 4-6 muestra, para cada participante que intervino en el experimento, el valor observado, el valor que se hubiera predicho utilizando el modelo de predicción, los errores (diferencias) y los errores cuadráticos.

En este ejemplo, la suma de los errores cuadráticos, al predecir utilizando el modelo de predicción (SS_{Error}), es de 39,65. Para calcular la reducción proporcional del error, también se necesita la suma del error cuadrático al predecir utilizando la media (SS_{Total}). El resultado es 0,72. (Si el alumno lo desea puede controlar el resultado calculándolo por sí mismo). Cabe recordar que, para obtener SS_{Total}, primero se debe calcular cada valor menos la media para obtener el error. Luego se eleva al cuadrado cada uno de esos errores y se suman. Por ejemplo, en el caso del primer participante, el error cuadrático al predecir utilizando la media es el valor 4 menos la media de 5,6, lo que da un error de – 1,6 y un error cuadrático de 2,56).

En este ejemplo, utilizar la norma de predicción reduce el error cuadrático casi a la mitad, de 72 a 39,65. Para ser precisos, al dividir la reducción de 32,35 por el SS_{Total} de 72, resulta en una reducción proporcional del error de 0,45 (ó 45%). El mismo es representado mediante la fórmula:

Reducción proporcional del error =

$$\frac{SS_{Total} - SS_{Error}}{SS_{Total}} = \frac{72 - 39{,}65}{72} = \frac{32{,}35}{72} = 0{,}45$$

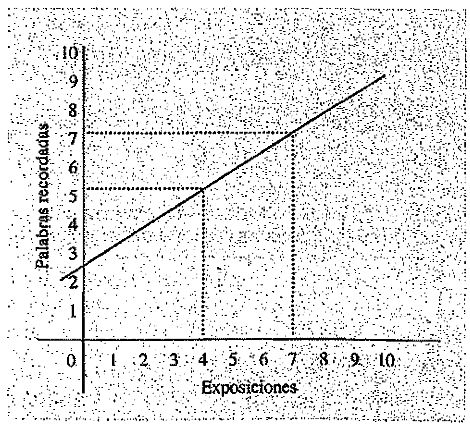


Figura 4-5. Recta de regresión del ejemplo de las palabras recordadas, en la que se indica la cantidad predicha de palabras recordadas por individuos que tuvieron cuatro y siete exposiciones de cada palabra.

Esta cifra también coincide (teniendo en cuenta los redondeos) con el cuadrado del coeficiente de correlación.

$$\text{Reducción proporcional de error} = r^2 = 0{,}68^2 = 0{,}46$$

Finalmente, la figura 4-6 muestra el diagrama de dispersión con la correspondiente recta de regresión.

Tabla 4-6.
Valores observados y predichos y errores en el experimento que indagan el efecto de la cantidad de exposiciones sobre la cantidad de palabras recordadas (datos ficticios).

Sujeto	Cantidad de exposiciones	Cantidad de palabras recordadas		Error	Error²
	X	Y	$\hat{Y}$		
1	1	4	3,4	0,6	0,36
2	1	3	3,4	–0,4	0,16
3	2	3	4,0	1,0	1,00
4	2	5	4,0	1,0	1,00
5	3	6	4,6	1,4	1,96
6	3	4	4,6	–0,6	0,36
7	4	4	5,3	–1,3	1,69
8	4	6	5,3	0,7	0,49
9	5	5	5,9	–0,9	0,81
10	5	7	5,9	1,1	1,21
11	6	2	6,5	–4,5	20,25
12	6	9	6,5	2,5	6,25
13	7	6	7,1	–1,1	1,21
14	7	8	7,1	0,9	0,81
15	8	9	7,8	–1,3	1,69
16	8	8	7,8	0,2	0,40
				$SS_{Error} =$	39,65

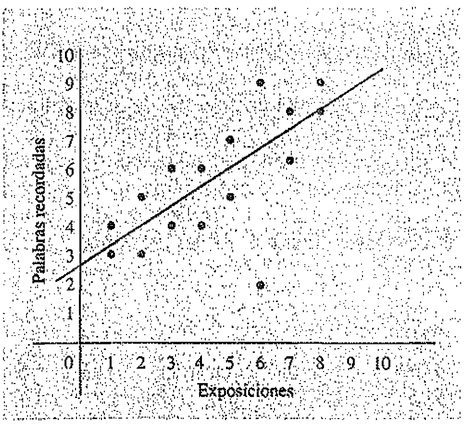


Figura 4-6. Diagrama de dispersión del ejemplo que trata sobre la cantidad de palabras recordadas, con la correspondiente recta de regresión.

EXTENSIÓN A CORRELACIÓN Y REGRESIÓN MÚLTIPLES

Hasta aquí hemos aprendido a predecir el valor de una persona en la variable dependiente utilizando el valor de esa misma persona en una sola variable predictora. Es decir, se predice una variable dependiente (como puede ser el nivel de estrés) sobre la base de una variable predictora (como la cantidad de personal supervisado). ¿Qué sucedería si se pudieran utilizar variables predictoras adicionales? Por ejemplo, al predecir el nivel de estrés de los gerentes, supongamos que además de la cantidad de personal supervisado, también se conociera el nivel de ruido y los plazos que cada gerente tiene que cumplir cada mes. Con esta información adicional, se podría realizar una predicción del nivel de estrés mucho más acertada.

La asociación entre una variable dependiente y dos o más variables se denomina **correlación múltiple**. Realizar predicciones en la situación anteriormente descripta se denomina **regresión múltiple**.[2]

Los detalles de la lógica y de los procedimientos de cálculo, y los puntos delicados de la utilización de la regresión múltiple exceden el alcance de un libro introductorio. Sin embargo, sí podemos presentar suficiente terminología e ideas clave como para que, al leer artículos de investigación, se pueda comprender el tema en forma general.

Es importante comprender los fundamentos de la regresión múltiple dado que la misma es muy común en la investigación psicológica. La regresión múltiple se utiliza prácticamente en uno de cada cinco artículos publicados en la revista científica de psicología social más importante (Reis & Stiller, 1992). Es probablemente más común aun en artículos de investigación de otras áreas de la psicología, como la del desarrollo, la personalidad, la psicología clínica y la mayoría de las áreas aplicadas. De hecho, la regresión múltiple es mucho más común que la regresión bivariada. En este capítulo, hemos enseñado la regresión bivariada con cierto detalle principalmente para crear los cimientos que permitan comprender este procedimiento más abstracto.

Modelo de predicción con puntuaciones Z para la regresión múltiple

En el caso de la regresión múltiple, cada variable predictora tiene su propio coeficiente de regresión. Para encontrar la puntuación Z predicha de la variable dependiente, se multiplica la puntuación Z de cada variable de predicción por su beta (coeficiente de regresión estandarizado) y luego se suman los productos obtenidos. En símbolos:

$$\hat{Z}_Y = (\beta_1)(Z_{X_1}) + (\beta_2)(Z_{X_2}) + (\beta_3)(Z_{X_3}) \tag{4-8}$$

En esta fórmula, β_1 es el coeficiene de regresión estandarizado de la primera variable predictora; β_2 y β_3 son los coeficientes de regresión estandarizados de la segunda y tercera variable. Z_{X_1} es la puntuación Z de la primera variable predictora; Z_{X_2} y Z_{X_3} son las puntuaciones Z de la segunda y tercera variables predictoras $(\beta_1)(Z_{X_1})$, significa multiplicar β_1 por Z_{X_1}, y así sucesivamente.

Más adelante se detalla el modelo de regresión múltiple correspondiente al ejemplo del nivel de estrés de los gerentes. No hemos enseñado cómo calcular los β_1 correspondientes, debido a que ese cálculo está muy lejos del alcance de un texto introductorio. Por otro lado, en las investigaciones, casi siempre se calculan por computadora. En este modelo, el nivel de estrés es *Y*, la

[2] También existen procedimientos que permiten utilizar más de una variable dependiente. Por ejemplo, podría ser necesario averiguar en qué medida la variable predictora "cantidad de empleados" supervisados es adecuada, tanto para el nivel de estrés como para la cantidad de ausentismo. Los procedimientos que involucran más de una variable dependiente se denominan de "estadística multivariada" y son bastante avanzados. En el capítulo 17 se presentan algunos ejemplos.

cantidad de empleados supervisados es X_1, el nivel de ruido es X_2 y la cantidad de plazos que se deben cumplir por mes es X_3.

$$\hat{Z}_Y = (0{,}51)(Z_{X_1}) + (0{,}11)(Z_{X_2}) + (0{,}33)(Z_{X_3})$$

Supongamos que se intenta predecir el nivel de estrés de un nuevo gerente que tenía una puntuación Z de 1,27 correspondiente a la cantidad de empleados para supervisar (una cantidad bastante alta), una puntuación Z de −1,81 con respecto al ruido en las condiciones de trabajo (un bajo nivel de ruido) y una puntuación Z de 0,94 en relación con la cantidad de plazos que se deben cumplir por mes (una cantidad un poco alta de vencimientos). Para encontrar la puntuación Z predicha del nivel de estrés, se debe multiplicar 0,51 por la puntuación Z de empleados supervisados, 0,11 por la puntuación Z de nivel de ruido y 0,33 por la puntuación Z de los vencimientos. Luego, se deben sumar los resultados.

$$\hat{Z}_Y = (0{,}51)(1{,}27) + (0{,}11)(-1{,}81) + (0{,}33)(0{,}94) = 0{,}65 + -0{,}20 + 0{,}31 = 0{,}76$$

Por lo tanto, para un gerente que trabaja en esas condiciones se predeciría una puntuación Z de nivel de estrés de 0,76. Es decir, un nivel de estrés de aproximadamente tres cuartas partes de un desvío estándar por sobre la media.

Relación entre los coeficientes beta de la regresión múltiple y las correlaciones comunes

Existe una diferencia particularmente importante entre la regresión múltiple y la predicción cuando se utiliza sólo una variable de predicción. En la regresión bivariada, $\beta = r$. En la regresión múltiple, en general β no es igual a r. Es decir, el beta de una variable predictora en particular **no** es igual a la correlación común de esa variable predictora con la variable dependiente. En la mayoría de los casos, beta será menor (más cercana a 0) que r.

La razón de esta discrepancia es que las variables predictoras generalmente están correlacionadas entre sí. Por lo tanto, parte de aquello que hace de una variable predictora un exitoso medio de predicción de la variable dependiente se superpone con lo que hace a las otras variables predictoras exitosas para predecir la variable dependiente. Por lo tanto, las correlaciones de cada variable predictora con la variable dependiente son, en cierta medida, redundantes, ya que lo que contiene cada variable de predicción se superpone con lo que contienen las otras variables predictoras. Sin embargo, esto no sucede con los beta. En la regresión múltiple, beta se calcula de modo que pueda ser la contribución única y distintiva de la variable predictora a la predicción de la variable dependiente. Los coeficientes beta excluyen cualquier superposición con otras variables de predicción.[3]

Analicemos el ejemplo del nivel de estrés de los gerentes. Cuando realizamos la predicción utilizando sólo la cantidad de empleados supervisados, beta era igual al coeficiente de correlación

[3] Técnicamente, la contribución única a la reducción proporcional del error de una variable predictora, en el contexto de las otras variables de predicción, es un cálculo estadístico denominado **correlación semiparcial cuadrática** (sr^2), un número que ocasionalmente aparece en artículos de investigación. Sin embargo, es más común que los investigadores de aspectos psicológicos informen sólo las betas y luego hablen de ellas como indicadores aproximados de la contribución única de una variable. Siempre que se tenga en cuenta que son "aproximados", esto resulta razonable, ya que beta y sr^2 están estrechamente relacionadas. Una beta alta generalmente corresponde a una sr^2 alta, el signo (positivo o negativo) de una beta es siempre el mismo que el de una sr^2, y la significación de una beta es siempre la misma que la de sr^2. En todo caso, debido a este uso común (y además porque tratar adecuadamente el tema de sr^2 excede el alcance de un texto introductorio), nuestra exposición adopta esta interpretación amplia de beta como indicador de la contribución única de una variable a la predicción.

de 0,88. Ahora bien, en el ejemplo con regresión múltiple, el beta de empleados supervisados es de sólo 0,51. Beta es menor debido a que parte de lo que hace que la cantidad de empleados supervisados pueda predecir el nivel de estrés se superpone con aquello que hace que el ruido y la cantidad de vencimientos predigan el estrés. (Por ejemplo, parte de lo que hace que la cantidad de personas supervisadas prediga el estrés es que esa cantidad de personas supervisadas aumenta el nivel de ruido).

Regresión múltiple con puntuaciones originales

Al igual que con la regresión bivariada, en la regresión múltiple es posible utilizar una fórmula de predicción con puntuaciones originales. Con tres variables de predicción y la constante de regresión a (si este concepto no está muy presente en la memoria del alumno, es conveniente que releyera la sección de regresión bivariada con puntuaciones originales), la fórmula es la siguiente:

$$\hat{Y} = a + (b_1)(X_1) + (b_2)(X_2) + (b_3)(X_3) \tag{4-9}$$

Supongamos que en el ejemplo del nivel de estrés de los gerentes contáramos con la información necesaria con respecto a los cuatro gerentes, y calculáramos los datos utilizando un procedimiento similar al que utilizamos para la regresión bivariada con puntuaciones originales. El modelo de predicción con puntuaciones originales podría ser el siguiente:

$$\hat{Y} = -4{,}70 + (0{,}56)(X_1) + (0{,}06)(X_2) + (0{,}86)(X_3)$$

Supongamos que un posible gerente iba a supervisar a 8 personas, con un muy alto nivel de ruido de 85 decibeles y con 4 vencimientos por mes *(el cual es mayor que el promedio de 3). El nivel de* estrés esperado sería bastante alto:

$$\begin{aligned}\hat{Y} &= -4{,}70 + (0{,}56)(8) + (0{,}06)(85) + (0{,}86)(4)\\ &= -4{,}70 + 4{,}48 + 5{,}1 + 3{,}44 = 8{,}32\end{aligned}$$

Es decir, la predicción del nivel de estrés de este gerente sería de 8,32.

Cada coeficiente de regresión de puntuaciones originales (b) es la razón de cambio de las puntuaciones originales de la variable predictora correspondiente, en el contexto de las otras variables de predicción. Así, a determinado nivel de cantidad de empleados supervisados y ruido, cada vencimiento adicional aumenta el valor de estrés predicho en 0,86 puntos. De modo similar, a cualquier nivel de empleados supervisados y cantidad de vencimientos, cada decibel de ruido aumenta el valor de estrés predicho en 0,06 puntos; a cualquier nivel de ruido y cantidad de vencimientos, cada persona adicional supervisada aumenta el valor de estrés predicho en 0,56 puntos.

Coeficiente de correlación múltiple

El **coeficiente de correlación múltiple** (generalmente simbolizado con R) describe la correlación total entre las variables predictoras, tomadas en su conjunto, y la variable dependiente.

Todas las variables predictoras contribuyen a la correlación. Por lo tanto, R debe ser al menos tan alta como la correlación individual más alta de una variable predictora con respecto a la variable dependiente. Sin embargo, por lo general cada variable predictora se superpone con las otras en su asociación con la variable dependiente. Por lo tanto, usualmente el coeficiente de correlación múltiple es menor que la suma de las correlaciones de cada variable predictora con la variable dependiente.

En el ejemplo del nivel de estrés de los gerentes, si suponemos que las tres r con respecto a la variable dependiente son de 0,88 (empleados supervisados con respecto al nivel de estrés), 0,38 (nivel de ruido con respecto al nivel de estrés) y 0,63 (vencimientos con respecto a nivel del estrés), la correlación múltiple de la variable de estrés con las tres predictoras tomadas en conjunto deberá ser al menos de 0,88 (la mayor de las tres correlaciones). Es decir, la correlación con las tres variables en su conjunto no podría ser menor que la correlación con cualquiera de ellas por separado. Por otro lado, R no podría ser mayor que la suma de los valores r, que en este caso es igual a 1,89 (0,88 + 0,38 + 0,63). De hecho, R, al igual que r, nunca podrá ser mayor a 1. No importa cuántas variables de predicción se utilicen. En el ejemplo, $R = 0{,}96$ (R tiene un rango de 0 a 1. A diferencia de r, R no puede ser negativa, por razones por las que no nos preocuparemos ahora. Cabe recordar, sin embargo, que beta o b, ya sea en la regresión bivariada o múltiple, pueden ser negativas y mayores que 1).

Reducción proporcional del error en la regresión múltiple

Finalmente, al igual que con la predicción bivariada, se puede calcular la reducción proporcional del error. Error en la regresión múltiple es lo mismo que error en la regresión bivariada, de modo que, si es necesario, es válido repasar la exposición que hemos hecho anteriormente sobre este tema. Como siempre, el error se calcula tomando el valor real y restándole el valor predicho. Sin embargo, en este caso, el valor predicho se obtiene utilizando un modelo de predicción de regresión múltiple. Una vez que se obtienen los valores predichos, el error cuadrático, la suma de los errores cuadráticos (SS_{Error}) y la reducción proporcional del error cuadrático, también se calculan todos exactamente del mismo modo que con la predicción bivariada. En la regresión múltiple, como en la regresión bivariada, la reducción proporcional del error compara SS_{Error} con SS_{Total} (SS_{Total}, cabe recordar, es la suma de los errores cuadráticos que surgen al utilizar la media de la variable dependiente como valor predicho para esa variable). Asimismo, al igual que con la predicción bivariada, el resultado es el cuadrado del coeficiente de correlación (en este caso, R^2). En el ejemplo, si $R = 0{,}96$, $R^2 = 0{,}92$.

Finalmente, al igual que con la predicción bivariada, R^2 también es la proporción de varianza explicada. Es decir, R^2 indica cuánto de la variación en la variable dependiente es explicada (predicha) por el conjunto de variables de predicción. En el ejemplo, el 92% de la variación en el nivel de estrés de los gerentes está explicada por la cantidad de empleados supervisados en el nivel de ruido y en la cantidad de vencimientos por mes.

Ejemplo de regresión y correlación múltiples

Watts y Wright (1990) entregaron cuestionarios sobre delincuencia y consumo de sustancias a estudiantes secundarios de sexo masculino y a delincuentes convictos del mismo sexo que residían en las instalaciones de la Comisión Juvenil de Texas. La tabla 4-7 muestra los resultados correspondientes a uno de los grupos étnicos estudiados. Como se puede observar, existen coeficientes de correlación considerables entre el nivel de delincuencia violenta y el consumo de cada tipo de sustancia. Sin embargo, es interesante observar lo que sucede cuando las variables predictoras se consideran en su conjunto (en la ecuación de regresión múltiple). Los coeficientes beta varían considerablemente. Planteado como un modelo de regresión múltiple con puntuaciones Z, sería:

Tabla 4-7.
Consumo de drogas como predictor de la delincuencia.

Droga consumida	r	β
Alcohol	0,415	–0,007
Tabaco	0,415	0,183
Marihuana	0,513	–0,046
Otras drogas ilegales	0,712	0,677

$R = 0{,}729$; $R^2 = 0{,}531$

Fuente: Watts, W., & Wright, L. (1990). "La relación entre el consumo de alcohol, tabaco, marihuana y otras drogas ilegales con la delincuencia entre adolescentes americanos-mexicanos, negros y blancos de sexo masculino". *Adolescencia*, 25, 171–181. Reimpreso con autorización.

$$Z_{\text{Delincuencia}} = (-0{,}007)(Z_{\text{Alcohol}}) + (0{,}183)(Z_{\text{Tabaco}}) + (-0{,}046)(Z_{\text{Marihuana}}) + (0{,}677)(Z_{\text{Otras}})$$

Al considerarlos en combinación, el principal factor al realizar predicciones de delincuencia violenta parecería ser el consumo de "otras drogas ilegales", y un factor secundario podría ser el consumo de tabaco. Es decir que conociendo el consumo de otras drogas y de tabaco, el conocimiento del consumo de alcohol y marihuana no agrega mucho a la capacidad para realizar predicciones sobre la delincuencia. Por ejemplo, la predicción sobre la base del consumo de alcohol es bastante importante si se lo considera individualmente (0,415), pero es casi insignificante (–0,007) cuando se conoce el consumo de otras drogas. Esto sucede porque toda información para la predicción, aportada por el conocimiento del nivel de consumo de alcohol, probablemente ya es aportada por el conocimiento del nivel de consumo de otras drogas por parte de esa persona.

(Cabe recordar que para este estudio se utilizó un diseño de correlación. Por lo tanto, no podemos estar seguros de cuál es la causa y cuál el efecto. Muy bien podría ser que el consumo de drogas fuera el resultado y no la causa de la delincuencia. También es posible que un tercer factor, como el tipo de ambiente en el que fueron criados los jóvenes, sea la causa tanto del consumo de sustancias como del nivel de delincuencia.

También se podría utilizar la fórmula de regresión múltiple para realizar predicciones. Supongamos que estuviéramos interesados en predecir el grado de delincuencia violenta de un joven con una puntuación Z de –1 en el consumo de alcohol, una Z de 0 (la media) con respecto al consumo de tabaco, una Z de +1 con respecto al consumo de marihuana y una Z de –2 con respecto al consumo de otras drogas ilegales. Utilizando el modelo de regresión múltiple con puntuaciones Z, la predicción sería la siguiente:

$$\begin{aligned} Z_{\text{Delincuencia}} &= (-0{,}007)(-1) + (0{,}183)(0) + (-0{,}046)(1) + (0{,}677)(-2) \\ &= 0{,}007 + 0 + -0{,}046 + -1{,}354 = -1{,}393 \end{aligned}$$

En el caso de este joven, se predeciría un registro bastante bajo de delincuencia violenta (1,393 desvíos estándares por debajo de la media).

Supongamos que otro joven tenía exactamente el mismo patrón, pero no con respecto a drogas ilegales, donde presentaba un alto nivel de consumo, digamos, una puntuación Z de + 2.

$$
\begin{aligned}
Z_{\text{Delincuencia}} &= (-0{,}007)(-1) + (0{,}183)(0) + (-0{,}046)(1) + (0{,}677)(2) \\
&= 0{,}007 + 0 + -0{,}046 + 1{,}354 = 1{,}315
\end{aligned}
$$

Para este joven se esperaría un alto registro de delincuencia violenta.

Finalmente, analicemos el caso de un joven con el mismo patrón que el primero (el que presentaba un bajo nivel de consumo de otras drogas ilegales y al que se le predijo una puntuación Z de −1,393), excepto que este joven fuma mucho, con una puntuación Z correspondiente a consumo de tabaco de +2.

$$
\begin{aligned}
Z_{\text{Delincuencia}} &= (-0{,}007)(-1) + (0{,}183)(2) + (-0{,}046)(1) + (0{,}677)(2) \\
&= 0{,}007 + 0{,}366 + -0{,}046 + -1{,}354 = -1{,}027
\end{aligned}
$$

Aunque parezca sorprendente, para este joven también se esperaría un nivel bajo de delincuencia violenta, ya que uno podría asociar con la delincuencia el hecho de fumar mucho. En la mayoría de los casos, los índices altos con respecto a fumar se asocian con la delincuencia, como lo indica el r de 0,415. Sin embargo, la gente joven que fuma generalmente también consume otras drogas ilegales (al menos así lo indica la información correspondiente a este grupo en este entorno particular), y esa parecería ser la razón por la cual, cuando se consideró separadamente el hecho de fumar, ésta estaba más fuertemente asociado con la delincuencia.

Es el momento de hacer un paréntesis. Esperamos que el ejemplo referido a predicciones sobre el potencial criminal de una persona, y su posible encarcelamiento, haya cambiado la percepción del lector. Cuando utilizamos grandes estudios para realizar predicciones sobre una sola persona, inmediatamente percibimos el estereotipo intrínseco y las posibles injusticias. No es de extrañarse que a menudo la gente desconfíe de la estadística. Pero ella es sólo una herramienta para analizar el futuro, como lo son la intuición o la experiencia clínica, y es tan compasiva como la persona que la utiliza para tomar una decisión. Si una persona insensible cita números "fríos" para justificar una decisión prejuiciosa, no son los números los que son fríos. (En el cuadro 4-1 ofrecemos un breve debate sobre el tema).

Antes de abandonar este estudio utilizado como muestra, será ilustrativo examinar el R correspondiente a esta información. El R de 0,729 es mayor que el r común más alto (que era 0,712). Sin embargo, R es en realidad considerablemente menor que la suma de los valores r individuales. (De hecho, la suma daría como resultado más de 1, lo cual, como mencionamos anteriormente es, como valor de R, imposible). Finalmente, R_2 es 0,531. Esto indica que si se realizaran predicciones utilizando este modelo de regresión múltiple para cada joven del grupo estudiado, el error cuadrático promedio en la predicción de los valores observados de delincuencia sería un 53,1% menor que si se utilizara la media de los valores de delincuencia como predictor de los valores individuales. En términos de proporción de varianza explicada, el 53,1% de la variación en la delincuencia de este grupo es explicada por las variables de consumo de drogas.

Cuadro 4-1.
Predicción clínica versus predicción estadística.

En 1954, Paul Meehl escribió un pequeño e inquietante libro titulado *Predicción estadística versus predicción clínica.* En él sostenía que cuando algunos expertos, tales como por ejemplo psicólogos clínicos (o gerentes de negocios, analistas económicos, ingenieros o médicos, entre otros), utilizan los tipos de procesos cognitivos internos no especificados, a los que comúnmente llamamos "intuiciones capacitadas", para realizar predicciones importantes y decisivas, no son, en líneas generales, ni remotamente tan precisos como podría serlo cualquier otro sujeto empleando **fórmulas** muy simples y directas. Por ejemplo, en el caso de la realización de un diagnóstico psiquiátrico, la entrevista y el diagnóstico de un clínico supuestamente bien capacitado son menos útiles que una simple regla, como lo es la del tipo utilizada en los procedimientos de regresión múltiple: "Si la persona ya ha ingresado al hospital dos veces, tiene más de 50, y aparentemente es suicida, entonces..."

Durante la primera década que prosiguió al cuestionamiento por parte de Meehl acerca de la precisión de los expertos, se realizaron considerables esfuerzos para refutarlo. Pero, en general, el descubrimiento de Meehl se ha mantenido (Dawes et al., 1993; Kleinmuntz, 1990): la cognición humana por sí sola es, en líneas generales, menos precisa al realizar predicciones que el método estadístico de análisis de regresión. Y se trata de predicciones importantes: nos referimos a los diagnósticos, a las que determinan una libertad condicional, o a decisiones comerciales y de ingeniería.

No debemos olvidar que hablamos de predicciones, por lo cual los métodos estadísticos tampoco son perfectos. Su ventaja principal es la coherencia, como un apostador que dispone de un sistema para jugar. Los seres humanos explicamos mejor por qué sucedió algo después de que sucedió, porque entonces sabemos dónde buscar la causa. Pero debido a que las predicciones a menudo tienen serias consecuencias, aún es desconcertante descubrir que la lógica o la intuición humana, después de largas entrevistas o pruebas, puedan ser tan poco eficientes comparadas con una simple fórmula.

Naturalmente, el centro de la atención se ha enfocado en el funcionamiento de la cognición, el por qué de su imperfección y qué puede hacerse para mejorarla, si es que algo puede hacerse. Su imperfección se debe principalmente a que las personas suelen realizar correlaciones ilusorias (*véase* cuadro 3-2) o son demasiado confiadas: no llevan un registro de sus éxitos y fracasos para controlar si en realidad son precisos, sino que dan demasiada importancia a los éxitos recordados y olvidan sus fracasos. Además, lamentablemente, el exceso de confianza proviene en parte de la experiencia, que en realidad rara vez ayuda demasiado, ya que no aporta información acerca del resultado de un proceso (los clínicos pueden realizar cientos de diagnósticos sin enterarse luego si estaban en lo cierto). Finalmente, la memoria y la cognición humana pueden no tener la capacidad de manejar la información, como tampoco las operaciones necesarias para tomar ciertas decisiones complejas.

Gran cantidad de investigaciones se han dedicado al tema de cómo "quitar el sesgo" a las decisiones humanas. Se puede mostrar a los profesionales en qué casos la intuición será más precisa (por ejemplo, cuando se necesita un trabajo rápido y por lo tanto no delicado, o cuando basta con un simple promedio) y cuándo es preferible utilizar una fórmula (cuando hay tiempo para la deliberación o cuando las reglas son más complicadas).

También existe una cantidad considerable de trabajos sobre los distintos instrumentos de ayuda para la toma de decisiones, como pueden ser los programas informáticos que incluyen reglas para la toma de decisiones, aportadas por los mismos expertos. En algunos casos, expertos bien informados sobre la situación particular pueden agregar más información intuitiva o subjetiva de último momento (Holzworth, 1996; Whitecotton, 1996). Aunque los mecanismos de ayuda para la toma de decisiones puedan parecer inflexibles, y por lo tanto "inhumanos", estos mecanismos y fórmulas pueden modificarse tantas veces como sea necesario. Lo que no debe hacerse con ninguno de estos mecanismos de ayuda es dejarlos de lado cada vez que una persona tiene un presentimiento que le indica que puede desempeñarse mejor sin ellos.

Sin embargo, habiendo resumido todo esto, Kleinmuntz (1990) observó que en la mayoría de los casos en los que es necesario tomar decisiones, aún se utilizan los dictámenes humanos en lugar de las fórmulas o fórmulas combinadas con cognición, que son más acertadas. Cuando las apuestas son altas, como en los casos de vida o muerte, la mayoría de las personas aún tiene más confianza en las decisiones humanas, tal vez por la esperanza de que la intuición inspirada pueda acertar en un caso en particular. Además, las personas creen, tal vez con razón, que los complejos patrones que presentan las situaciones reales son captados mejor por los expertos más allegados y acostumbrados a esas situaciones. En tercer lugar, las fórmulas para la toma de decisiones no existen o no están al alcance de las personas que las necesitan. Finalmente, el costo de la creación y prueba de una fórmula para la toma de decisiones es, a menudo, prohibitivo.

Aun así, la utilización de "sistemas de apoyo para la toma de decisiones" está en crecimiento. Por ejemplo, jugadores expertos de ajedrez han desarrollado sistemas de ayuda que algunas veces pueden ser más "inteligentes" que sus propios creadores, por el simple hecho de ser completamente coherentes. Es así como algunos jugadores de ajedrez se sienten cómodos utilizando sistemas de apoyo para la toma de decisiones, con el fin de mantener una línea durante el juego. Es lógico esperar que médicos, psicólogos clínicos e ingenieros también adopten fórmulas con reglas generadas por ellos mismos, particularmente para contrarrestar los efectos del cansancio o el interés emocional. Posiblemente lleve tiempo, pero todos tendremos que reconocer que puede ser más humano evitar decisiones subjetivas y preocupantes cuando existen sistemas de apoyo objetivos.

Otro ejemplo

Analicemos otro ejemplo. Terpstra y Rozell (1997) realizaron un estudio sobre la manera en que los directores de personal de empresas obtienen información sobre nuevos desarrollos en su campo. Los investigadores enviaron cuestionarios a una muestra de grandes empresas norteamericanas elegidas al azar. Los cuestionarios solicitaban a los gerentes de personal que indicaran en qué medida utilizaban distintas fuentes de información, incluidas fuentes académicas (tales como artículos de investigación), fuentes profesionales (tales como revistas profesionales de comercio) y consultores profesionales. Los cuestionarios también indagaban sobre la rentabilidad de la empresa durante los últimos cinco años.

La tabla 4-8 muestra los resultados, divididos por tipo de empresa. Examinemos los resultados de las empresas de servicio. El R^2 total era de 0,60, es decir, que al predecir la rentabilidad se puede reducir un 60% del error cuadrático, conociendo la medida en la que se utilizan estas distintas fuentes de información. Ahora veamos las correlaciones bivariadas. Queda claro que las fuentes académicas son muy importantes. La correlación entre las fuentes académicas y la rentabilidad era de 0,64, una correlación bastante importante. La utilización de fuentes profesionales también estaba fuertemente relacionada con la rentabilidad, mientras que la utilización de consultores como fuente de información tenía una correlación con la rentabilidad de sólo el 0,23.

Pasemos ahora al coeficiente de regresión. Cabe destacar que son *b* y no beta, y que no están estandarizados, aunque en su artículo los investigadores aclaran que antes de realizar el análisis de regresión múltiple convirtieron los valores de utilización de información en puntuaciones Z. Por lo tanto, las variables predictoras, siendo todas puntuaciones Z, están en la misma escala. Es decir que las diferencias entre los *b* no se deben a que las variables predictoras estén en diferentes escalas, sino a las diferencias entre las asociaciones particulares de cada *b* con la variable dependiente.

La principal cuestión que se desprende de la tabla es que las fuentes académicas y profesionales presentaban las relaciones particulares más importantes con respecto a la rentabilidad.

Sin embargo, examinar las cifras referidas a consultoría resultará especialmente interesante con respecto a lo que significa la regresión múltiple. La consultoría presentó una correlación bivariada positiva de 0,23 con la rentabilidad. Aun así, en el contexto de la regresión múltiple, la relación de la consultoría con la rentabilidad es bastante negativa.[4] Es decir, considerada por sí sola, una mayor consultoría está ligada a una mayor rentabilidad. ¡Pero si tenemos en cuenta *cualquier nivel fijo de información académica y profesional, a mayor consultoría, menor rentabilidad*!

Una explicación posible para esta aparente paradoja sería que el valor positivo de la consultoría se superpone con la obtención de información de otras fuentes. Tal vez las empresas interesadas en obtener información utilizan más todas las fuentes. Por lo tanto, la correlación positiva entre consultoría y rentabilidad se debería a que ambas son causadas por un tercer factor (interés en la información en general). En realidad, una vez que se toma en cuenta esta tendencia general a obtener información, la consultoría podría dañar la rentabilidad debido a que es muy costosa. Esta es sólo una explicación posible. Lo importante es que aplicar la regresión múltiple reveló un patrón de resultados que podría no haberse notado antes de la investigación y que debería generar ideas con una nueva orientación.

Otro tema importante es que, al menos para el caso de las empresas de servicio, contar con un gerente de personal que lee artículos de investigación académica puede ser muy rentable.

[4] Este es un ejemplo de lo que técnicamente se denomina **supresión**. En el ejemplo que estamos analizando, puede considerarse que la asociación positiva general con la rentabilidad presenta dos aspectos: una asociación positiva y una asociación negativa. En este caso, el aspecto positivo se superpone con las otras variables de predicción. Por lo tanto, cuando se incluyen las otras variables de predicción en la regresión, se "suprime" la superposición. (Es decir, su influencia es eliminada de la asociación única entre la consultoría y la rentabilidad, representada por beta). El resultado es que sólo el aspecto restante, la asociación negativa, forma parte de beta. En términos más generales, la supresión ocurre siempre que el coeficiente de regresión de determinada variable de predicción es de signo opuesto a su correlación bivariada con la variable dependiente. (Existe otro tipo de situación que también presenta supresión: cuando beta es mayor que la correlación bivariada).

Tabla 4-8.
Resultados del análisis de regresión y correlación de la relación entre utilización de fuentes de información y rentabilidad en distintos tipos de empresas.

	Producción		Servicios		Venta por mayor/menor		Financieras	
Fuente de información	*b*	*r*	*b*	*r*	*b*	*r*	*b*	*r*
Académico	−0,09	0,04	0,72**	0,64**	−0,08	−0,07	0,17	0,26
Profesional	−0,05	0,09	0,45**	0,49**	0,08	−0,01	0,10	0,23
Consultoría	0,29*	0,22*	−0,36	0,23	−0,12	−0,11	0,12	0,06
R^2	0,06		0,60		0,02		0,10	
F	1,17		6,61**		0,11		0,37	

Nota: Los tamaños de las muestras, en el caso de empresas de producción, servicios, venta por mayor o menor y financieras eran de 63 a 65, 16 a 18, 20 a 22 y 13 a 15, respectivamente. No se realizaron análisis en el área de transporte o comunicación, como tampoco en las áreas de agricultura, minería o construcción, debido a restricciones con respecto al tamaño de las muestras.

*$p < 0,10$. **$p < 0,05$.

Fuente: Terpstra, D. E., & Rozell, E. J. (1997), tab 6. "Fuentes de información para recursos humanos y su relación con la rentabilidad institucional". *Periódico sobre Ciencia del Comportamiento Aplicada [Journal of Applied Behavioral Science]*, 33, 66–83. Copyright, 1997, por el NTL Institute, Inc. Reimpreso con autorización de Sage Publications. Inc.

CONTROVERSIAS Y LIMITACIONES

Todas las limitaciones que se plantearon al tratar el tema de la correlación (capítulo 3) se aplican en igual o mayor medida a la regresión bivariada y múltiple. Los cálculos de regresión subestiman el grado de posibilidad de predicción si la relación implícita es curvilínea, si el grupo estudiado tiene un rango restringido o si las medidas no son perfectamente confiables. Es decir, en cada uno de estos casos, R Y R^2 (y generalmente b y β) son menores de lo que deberían ser para reflejar el verdadero grado de asociación de las variables de predicción con la variable dependiente. La regresión por sí sola tampoco indica la dirección de causalidad implícita. La dirección de causalidad depende del diseño experimental (*véase* apéndice A). Es importante ser muy cuidadoso al leer artículos de investigación ya que, incluso en las publicaciones, a veces los investigadores pasan por alto estas limitaciones cuando analizan los resultados de regresiones complejas.

Existe actualmente una controversia con respecto a la regresión múltiple que cuestiona cómo juzgar la importancia relativa de las diferentes variables de predicción al predecir la variable dependiente. En cuanto a los fines de predicción exclusivamente, los coeficientes de regresión (tanto estandarizados como de puntuaciones originales) cumplen bien esa función, pero no necesariamente son ideales para comprender la importancia de los diferentes elementos de predicción desde el punto de vista teórico. Como observamos anteriormente, un coeficiente de regresión indica la contribución particular de la variable predictora a la predicción, independientemente de los otros predictores. Una variable puede tener aparentemente una importancia bastante diferente en relación con los otros predictores, cuando se predice sólo a partir de ella, sin tener en cuenta esos otros elementos (es decir, utilizando la correlación ordinaria entre esa variable y la variable dependiente). Por ejemplo, en el estudio de la delincuencia y el consumo de drogas, los coeficientes beta sugerían que el consumo de tabaco era más importante en la predicción de la delincuencia que el consumo de marihuana, pero las correlaciones ordinarias sugerían exactamente lo contrario. Más aún, si se agregaran otras variables de predicción, como el consumo de otras drogas ilegales, todo el patrón de coeficientes beta podría volver a cambiar. ¿Qué importancia se le atribuye entonces a una variable de predicción que muestra tantas facetas diferentes en tantos contextos diferentes?

El problema surge en la regresión múltiple debido a que las variables predictoras están correlacionadas entre sí. Esta situación se denomina **multicolinealidad**, y en cierto grado casi siempre está presente en la regresión múltiple. Por lo tanto, es sorprendente que no exista un método acordado sobre cómo juzgar la importancia relativa de las variables predictores. La falta de consenso no se debe a la falta de propuestas: a lo largo de los años se han estudiado una gran cantidad de métodos para solucionar este problema (*véase* Cohen & Cohen, 1983). La mayoría de los expertos recomiendan utilizar toda la información disponible acerca de los distintos aspectos de importancia relativa. Es decir, tener en cuenta tanto las correlaciones ordinarias como los coeficientes de regresión sin olvidar la diferencia entre lo que cada uno de estos datos indica. El coeficiente de correlación indica la asociación general de la variable predictora con la variable dependiente, mientras que el coeficiente de regresión indica la asociación individual de la variable predictora con la variable dependiente, más allá de las otras variables de predicción.

Además de estas y otras controversias relacionadas con los aspectos estadísticos, durante muchos años ha existido una controversia que actualmente continúa en vigencia y que se refiere a la superioridad de la predicción estadística con respecto a métodos más intuitivos, humanistas o clínicos. En el cuadro 4-1 se plantea este tema.

LOS MODELOS DE PREDICCIÓN SEGÚN SE DESCRIBEN EN ARTÍCULOS DE INVESTIGACIÓN

No es muy frecuente que los modelos de predicción bivariadas sean citados en artículos de investigación psicológica; en la mayoría de los casos se informan las correlaciones simples. Algunas veces se publican rectas de regresión de predicciones bivariadas. Esto ocurre generalmente cuando existe más de un grupo y el investigador quiere ilustrar la diferencia en la norma de predicción entre los dos grupos. Por ejemplo, analicemos un experimento realizado por Nezlek et al. (1997).

En ese experimento, los participantes escribieron descripciones de sí mismos y las intercambiaron con otros cuatro alumnos que también participaban del estudio. Luego, cada alumno clasificó en forma privada a los otros alumnos con respecto a cuánto les gustaría trabajar con ellos en la siguiente tarea. Se explicó que sólo tres de los cinco trabajarían juntos en dicha tarea. Luego, el investigador informó a la mitad de los participantes que habían sido seleccionados para trabajar en la siguiente tarea con el resto del grupo: ésta era la condición de inclusión. Los investigadores dijeron al resto de los participantes que no habían sido escogidos para trabajar con los otros y que trabajarían solos: la condición de exclusión. (En realidad los investigadores decidieron al azar, hecho que fue cuidadosamente explicado a todos los participantes cuando finalizó el estudio, para que nadie se sintiera mal).

Llegados a este punto, como parte de toda una serie de cuestionarios se preguntaba a los participantes cuán aceptados se sentían. Previamente, al comienzo del estudio, los participantes habían completado una escala de autoestima. La figura 4-7 muestra las líneas de regresión de los dos grupos experimentales. Cada recta de regresión indica qué grado del nivel de autoestima predijo sentimientos de aceptación. Se puede observar que en el caso de los alumnos en condición de exclusión, existía una relación muy marcada entre autoestima y aceptación. Aquellos con alta autoestima se sentían aceptados, aquellos con baja autoestima no se sentían para nada aceptados.

Sin embargo, en el grupo en condición de inclusión, la autoestima influyó mucho menos en el sentimiento de aceptación: todos se sentían bastante bien aceptados.

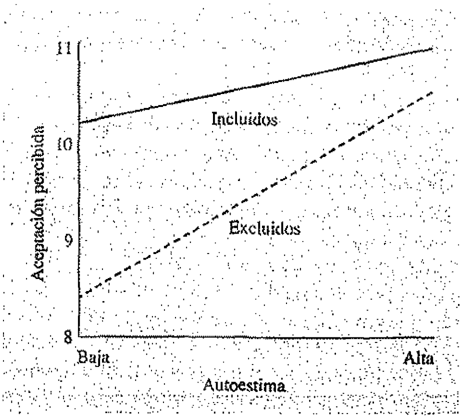


Figura 4-7. Efectos de la inclusión/exclusión y de la autoestima en la aceptación percibida. (**Fuente:** J. B., Kowalski, R. M., Leary, M. R., Blevins, T., & Holgate, S. (1997), fig. 1. "Características de la personalidad que moderan las reacciones al rechazo interpersonal: depresión y autoestima". *Boletín de Psicología Social y Personalidad [Personality and Social Psychology Bulletin]*, 23, 1235–1244.)

Como observamos anteriormente, los resultados de la regresión múltiple son comunes en los artículos de investigación, y a menudo se hace referencia a ellos en las tablas. Ya hemos visto algunos ejemplos (tablas 4-7 y 4-8). Frecuentemente, las tablas incluirán algunos otros cálculos estadísticos, además de aquellos que hemos tratado. Algunos están relacionados con la significación estadística (*véase* en el capítulo 3 una breve exposición sobre la significación del coeficiente de correlación); otros serán tratados en el capítulo 17. De todos modos, es posible comprender casi toda la información importante incluida en esas tablas sólo con lo aprendido aquí.

Analicemos los resultados de un estudio realizado por Jehn y Shah (1997) sobre el desempeño de grupos formados por tres personas que debían realizar en forma conjunta tareas físicas y de toma de decisiones en una situación de laboratorio. Los investigadores grabaron las interacciones en video y analizaron las cintas para estudiar varios aspectos de la interacción grupal. La tabla 4-9 muestra el coeficiente de regresión para la predicción del desempeño a partir de varias cualidades de interacción. Se puede observar que la comunicación positiva y la planificación presentan coeficientes beta relativamente bajos (y negativos), mientras que el compromiso, el control y la

Tabla 4-9.
Resumen del análisis de regresión con variables de predicción del desempeño.

Variable	*B*	*SE B*	β
Comunicación positiva	0,288	0,228	–0,127
Planificación	0,062	0,055	–0,190
Compromiso	1,340	0,134	0,432*
Control	1,210	0,049	0,449*
Cooperación	0,780	0,154	0,376*

Nota: N = 106, R = 0,55.
*$p < 0{,}01$.
Fuente: Jehn, K. A., & Shah, P. P. (1997), tab. 4. "Relaciones interpersonales y desempeño en las tareas: análisis de los procesos en grupos de amigos y conocidos". *Periódico de Psicología Social y Personalidad, [Journal of Personality and Social Psychology,]* 72, 775–790. Copyright, 1997, por la Asociación Americana de Psicología [American Psychological Association] Reimpreso con autorización.

cooperación fueron predictores del desempeño mucho más importantes. Además, se observa (en la parte inferior de la tabla) que la correlación total de los cinco predictores con el desempeño presentaba un R de 0,55, lo que constituye un dato sustancial.

Por otro lado, significa que a partir de estas cinco variables se explicó menos de un tercio de la varianza total del desempeño (es decir, $R^2 = 0{,}30$). La tabla también incluye, además de R y de coeficientes de regresión estandarizados y no estandarizados (rotulados aquí con B mayúscula), otros cálculos estadísticos: el error estándar (SE B) de cada B. El error estándar está relacionado con la precisión en la estimación de los coeficientes con respecto a la población en general. Este concepto será más fácil de comprender después de haber leído el capítulo 7.

RESUMEN

La predicción (o regresión) bivariada se utiliza para predecir valores de una variable dependiente sobre la base de valores de una variable predictora. La mejor norma o modelo para predecir la puntuación Z de una persona en una variable dependiente es multiplicar un número denominado coeficiente de regresión estandarizado (beta) por la puntuación Z de esa persona en la variable predictora. El mejor número para utilizar como coeficiente de regresión estandarizado en la predicción bivariada es el coeficiente de correlación.

También se pueden realizar predicciones con puntuaciones originales convirtiendo el valor observado de una persona en la variable predictora en la puntuación Z correspondiente, multiplicándolo por beta, y luego convirtiendo la resultante puntuación Z predicha de la variable dependiente nuevamente en una puntuación bruta. Los tres pasos anteriores pueden combinarse en una sola fórmula que permite predecir la puntuación original de una persona en la variable dependiente, a partir, directamente, de la puntuación original de esa persona en la variable predictora. Esta fórmula presenta dos partes principales: un coeficiente de regresión (denominado b) que se multiplica por la puntuación original de la persona en la variable dependiente y una constante de regresión (denominada a) que se suma al resultado. Si en un gráfico con las dos variables se dibujan los valores predichos, a través de esta fórmula para la variable dependiente se trazará la recta de regresión. La pendiente de la recta de regresión es igual al coeficiente de regresión para las puntuaciones originales; la constante de regresión indica dónde esta recta cruza el eje vertical (es la ordenada del punto de la recta con abscisa 0).

La exactitud de la predicción puede estimarse aplicando el modelo de predicción a los valores en los que se basó la correlación original. La diferencia entre cada valor observado y lo que hubiera sido predicho para ese individuo, utilizando el modelo de predicción, se denomina error. Elevando estos errores al cuadrado y sumándolos obtenemos la suma de errores cuadráticos (SS_{Error}). Luego, se compara SS_{Error} con la suma de errores cuadráticos obtenida utilizando sólo la media de la variable dependiente como valor predicho (SS_{Total}). La reducción del error cuadrático lograda utilizando el modelo ($SS_{\text{Total}} - SS_{\text{Error}}$), dividida por el error cuadrático al predecir utilizando la media de la variable dependiente (SS_{Total}), se denomina reducción proporcional de error o proporción de la varianza explicada, que es igual al cuadrado del coeficiente de correlación.

En la regresión múltiple, se predice una variable dependiente utilizando dos o más variables predictoras. Cada variable predictora se multiplica por su propio coeficiente de regresión, y los resultados se suman para realizar la predicción. (Cuando se utilizan puntuaciones originales, también se suma una constante de regresión). Cada coeficiente de regresión indica la relación del predictor con la variable dependiente en el contexto de las otras variables de predicción. El coeficiente de correlación múltiple describe el grado general de asociación entre la variable dependiente y las variables de predicción tomadas en su conjunto.

Las regresiones bivariada y múltiple tienen las mismas limitaciones que la correlación ordinaria. Además, en la regresión múltiple generalmente existe una ambigüedad considerable al interpretar la importancia relativa de las variables predictoras.

Términos clave

- Predicción bivariada.
- Regresión bivariada.
- Error.
- Correlación múltiple.
- Coeficiente de correlación múltiple (R).
- Regresión múltiple.
- Modelo de predicción.
- Proporción de varianza explicada (r^2, R^2).
- Reducción proporcional del error(r^2, R^2).
- Fórmula de predicción con puntuaciones originales.
- Coeficiente de regresión para puntuaciones originales (b).
- Coeficiente de regresión.
- Constante de regresión (a).
- Recta de regresión.
- Pendiente.
- Coeficiente de regresión estandarizado (b).
- Suma de los errores cuadráticos (SS_{Error}).
- Error cuadrático total al predecir utilizando la media (SS_{Total}).

Ejercicios

Los ejercicios implican la realización de cálculos (con la ayuda de una calculadora). La mayoría de los problemas estadísticos reales se resuelve por computadora. Pero aunque exista la posibilidad de utilizar una computadora, es conveniente realizar estos ejercicios a mano para incorporar el método de trabajo.

Para adquirir práctica en la utilización de una computadora, para resolver problemas estadísticos, se puede utilizar la sección de computación de cada capítulo, publicada en la *Guía de estudio y libro de tareas de computación para el alumno [Student's Study Guide and Computer Workbook]* que acompaña este texto.

Todos los datos de esta sección son ficticios (a menos que se especifique lo contrario).

Las respuestas a los ejercicios de la serie I se encuentran al final del libro.

SERIE I

1. Un psicólogo especializado en deportes, que trabaja con atletas de un deporte en particular, ha descubierto que los valores observados en una prueba de conocimientos sobre fisiología presentan una relación de 0,4 con la cantidad de lesiones sufridas durante el año subsiguiente. Ahora el psicólogo planea probar atletas nuevos y utilizar esta información para predecir la cantidad de lesiones que pueden llegar a sufrir. a) Indique la variable predictora, la variable dependiente y beta; b) escriba el modelo de predicción con puntuaciones Z, y c) indique puntuaciones Z predichas para la cantidad de lesiones que sufrirán los atletas cuyas puntuaciones Z en la prueba sobre fisiología son −2, −1, 0, +1 y +2.

2. Determine el modelo de predicción con puntuaciones originales para los puntos (a) a (g) que aparecen a continuación. Construya después un sólo gráfico que muestre todas las rectas de regresión, rotulando cada una con su letra correspondiente. (Construya un gráfico lo

suficientemente grande como para que las rectas queden *claramente separadas*).

	Variable dependiente (Y)		Variable de predicción (X)		
	M	SD	M	SD	r
(a)	10	2,0	10	2,0	0,4
(b)	20	2,0	10	2,0	0,4
(c)	10	2,0	20	2,0	0,4
(d)	10	2,0	10	4,0	0,4
(e)	10	4,0	10	2,0	0,4
(f)	10	2,0	10	2,0	–0,4
(g)	10	2,0	10	2,0	0,8

3. Un profesor ha descubierto que las notas en el examen parcial predicen las notas en el final. La fórmula de predicción con puntuaciones originales es:

Nota en el final = 40 + (0,5)(nota en el parcial)

Calcule las notas predichas para el examen final de cada uno de ocho alumnos, cuyas notas en el parcial fueron 30, 40, 50, 60, 70, 80, 90 y 100.

4. En el capítulo 3, serie I, ejercicio 1, describimos un estudio en el cual un investigador estaba interesado en la relación entre el grado de empatía que lograban los psicoterapeutas y el grado de satisfacción de sus pacientes con la terapia. Como estudio piloto, se analizaron cuatro parejas de pacientes y terapeutas. Más abajo se detallan los resultados, incluso las medias y los desvíos estándares. El coeficiente de correlación era 0,90, y el SS_{Total} correspondiente al grado de satisfacción del paciente era 10.

Número de pareja	Empatía terapeuta (X)	Satisfacción paciente (Y)
1	70,58	4,58
2	94,58	5,58
3	36,58	2,58
4	48,58	1,58
M	62	3
SD	22,14	1,58

a) Determine la fórmula de predicción con puntuaciones originales para predecir la satisfacción a partir de la empatía; b) utilice esta fórmula para encontrar los valores de satisfacción predichos para cada una de las cuatro parejas paciente-terapeuta; c) dibuje el diagrama de dispersión e incluya en él la recta de regresión; d) calcule el error y el error cuadrático para cada una de las cuatro predicciones; e) encuentre la reducción proporcional del error (utilizando SS_{Error} y SS_{Total}); f) halle la raíz cuadrada de la reducción proporcional del error calculada para comprobar si concuerda con el coeficiente de correlación, y g) explique los procedimientos realizados a alguien que comprende qué es la media, el desvío estándar, las puntuaciones Z y el coeficiente de correlación, pero que no sabe nada más sobre estadística.

5. En el capítulo 3, el ejercicio 2 de la serie I planteaba el caso de un instructor que preguntó a cinco estudiantes cuántas horas habían estudiado para un examen. Aquí mostramos la cantidad de horas de estudio y las calificaciones, junto con las medias y los desvíos estándares. La correlación era de 0,84 y la SS_{Total} correspondiente a las calificaciones era de 1.110. a) Determine la fórmula de predicción con puntuaciones originales para predecir las calificaciones a partir de las horas de estudio; b) utilice la fórmula para encontrar las calificaciones predichas para cada uno de los cinco estudiantes; c) dibuje el diagrama de dispersión e incluya en él la recta de regresión; d) calcule el error y el error cuadrático para cada una de las cinco predicciones; e) determine la reducción proporcional del error (utilizando SS_{Error} y SS_{Total}); f) saque la raíz cuadrada de la reducción proporcional del error calculada para comprobar si concuerda con el coeficiente de correlación, y g) explique los procedimientos realizados a alguien que comprende qué es la media, el desvío estándar, las puntuaciones Z y el coeficiente de correlación, pero que no sabe nada más sobre estadística.

	Horas de estudio (X)	Calificaciones (Y)
	0	52
	10	95
	6	83
	8	71
	6	64
M	6	73
SD	3,35	14,90

6. Interesados en la influencia que podría ejercer el estilo con que una madre ayuda a su hijo a comprender las interacciones sociales sobre la vida social real del niño, Mize y Pettit (1997) realizaron los arreglos necesarios para filmar en video a 43 madres voluntarias y a sus hijos de 3 a 5 años de edad, en tres sesiones independientes. En la sesión principal, se mostraban a las madres y a los niños cintas de video de otros niños que se comportaban de modo hostil o se rechazaban unos a otros; después, las madres discutían con los niños lo observado en los videos. Luego, los psicólogos clasificaban a cada madre según el "entrenamiento social", como por ejemplo, el modo en el que las madres habían ayudado a sus hijos a comprender lo que habían visto y les habían sugerido formas más positivas de manejar la situación. Se clasificaron los videos de las madres y los niños jugando según el "estilo de reacción" de las madres, es decir, la calidez y la capacidad de crear armonía con los niños. Finalmente, en la última sesión, se clasificaron los videos de los niños armando crucigramas en cuanto a la enseñanza "no social" por parte de las madres, es decir, el modo en que las madres ayudaban a sus hijos a desarrollar su capacidad de resolución de problemas. En otra etapa del estudio, los investigadores realizaron preguntas a los niños sobre cuánto les gustaban los otros niños. Utilizando esta información, pudieron obtener una medida general de cuánta apreciación gozaba cada niño, a lo que denominaron "aceptación por parte de sus pares".

Los investigadores desarrollaron la hipótesis de que se podría predecir la aceptación de un niño por parte de sus pares a partir de lo adecuada o inadecuada que fuera la madre como entrenadora social. También desarrollaron la hipótesis de que la relación entre el nivel de entrenadora social de la madre y la aceptación por parte de los pares se sostendría aun en una ecuación de regresión múltiple que incluyera entrenamiento no social, y en una ecuación de regresión que incluyera estilo de reacción.

La sección **"aceptación por sus pares"** de la tabla 4-10 muestra los resultados. La "ecuación 1" se refiere al modelo de regresión múltiple en el que se incluyen la enseñanza no social y el entrenamiento social como predictores de la aceptación por parte de pares. La "ecuación 2" se refiere al modelo de regresión múltiple en el que el estilo de reacción y el entrenamiento social se incluyen predictores de la aceptación por parte de los pares. Explique el significado de los resultados de aceptación por los pares como si se estuviera escribiendo para una persona que comprende qué es una correlación pero que nunca ha oído hablar de análisis de regresión o regresión múltiple. (Se puede ignorar la columna sr_1, "correlación semiparcial", *véase* nota al pie número 3. Todos los datos necesarios para interpretar esta tabla se encuentran en las columnas r, R^2 y beta).

7. a) Sobre la base de la tabla 4-10, sección "aceptación por los pares", escriba la ecuación 1 (una ecuación de regresión con puntuaciones Z). Luego calcule la puntuación Z predicha para la aceptación por los pares, correspondiente a niños cuyas madres presentan las siguientes puntuaciones Z.

Madre	Enseñanza no social	Entrenamiento social
A	−2	0
B	0	0
C	2	0
D	0	−2
E	0	2
F	2	2
G	−1	−2

b) Escriba la ecuación 2 y calcule la puntuación Z predicha para la "aceptación por los pares", correspondiente a niños cuyas madres presentan los siguientes puntuaciones Z:

Madre	Estilo de reacción	Entrenamiento social
A	−2	0
B	0	0
C	2	0
D	0	−2
E	0	2
F	2	2
G	−1	−2

Tabla 4-10.
Análisis de regresión simultáneo para la predicción de la habilidad social, la agresión y la aceptación por los pares, clasificados por maestros en el estudio 1.

	Criterio											
	Aceptación por pares				Habilidad social				Agresión			
Variables de predicción	r	R^2	sr_1	Beta	r	R^2	sr_1	Beta	r	R^2	sr_1	Beta
Ecuación 1:												
Enseñanza no social	0,21		0,10	0,10	0,15		0,05	0,06	–0,35*		–0,23	–0,24
Entrenamiento social	0,36*	0,14	0,30	0,32	0,31*	0,10	0,28	0,29	–0,41***	0,22**	–0,32	–0,33*
Ecuación 2:												
Estilo de recreación	0,34*		0,26	0,27	0,25		0,18	0,18	–0,26		–0,16	–0,17
Entrenamiento social	0,36*	0,19*	0,28	0,29	0,31*	0,13	0,25	0,26	–0,41***	0,20*	–0,36	–0,37*

Nota: sr_1 = correlación semiparcial; $n = 38$.
*$p < 0,10$; **$p < 0,05$; ***$p < 0,01$.

Fuente: Mize, J., & Pettit, G. S. (1997), tab. 2. "Entrenamiento social brindado por las madres, estilo de relación madre-hijo, y competencia de los niños con sus pares: ¿El medio es el mensaje?" *Desarrollo Infantil*, 68, 312–332. Copyright, 1997, por la Sociedad de Investigación del Desarrollo Infantil [Sociey for Research in Child Development] Inc. Reimpreso con autorización.

SERIE II

1. Elija algo que resulte interesante predecir y busque la información necesaria para poder predecirlo. (Ambas deberían ser cosas que puedan medirse en una escala numérica). Luego escriba el modelo de predicción, anotando el nombre de la variable predictora y el nombre de la variable dependiente. Además, estime un número para beta que tenga sentido, teniendo en cuenta lo aprendido sobre los valores que se están prediciendo. Finalmente, explique por qué se eligió ese tamaño de beta.

2. Determine el modelo de predicción con puntuaciones Z y el modelo de predicción con puntuaciones originales para cada uno de los siguientes casos. Además, prepare un solo gráfico que muestre todas las rectas de regresión (puntuaciones originales) y rotule cada una de ellas con la letra correspondiente desde la (a) hasta la (e), y que sea lo suficientemente grande como para que las rectas estén claramente separadas.

	Variable dependiente (Y)		Variable de predicción (X)		
	M	SD	M	SD	r
(a)	0	1,0	0	1,0	0,3
(b)	5	1,0	5	1,0	0,3
(c)	0	5,0	0	5,0	0,3
(d)	0	1,0	5	5,0	0,3
(e)	0	1,0	0	1,0	0,0

3. En el capítulo 3, serie II, ejercicio 1, cuatro individuos recibieron una prueba de destreza manual (valores altos significan mayor destreza) y una prueba de ansiedad (valores altos significan mayor ansiedad). A continuación indicamos los valores observados, medias y desvíos estándares. Calcule primero la correlación entre destreza y ansiedad (o refiérase a la respuesta en el capítulo 3). La SS_{Total} correspondiente a ansiedad era 84.

a) Determine la fórmula de predicción con puntuaciones originales para predecir la ansiedad a partir de la destreza; b) utilice la fórmula para calcular los valores de ansiedad predichos para cada uno de los cuatro individuos estudiados; c) dibuje el diagrama de dispersión e incluya en él la recta de regresión; d) calcule el error y el error cuadrático para cada una de las cuatro predicciones; e) calcule la reducción proporcional del error (utilizando SS_{Error} y SS_{Total}); f) saque la raíz cuadrada de la reducción proporcional del error calculada para controlar si concuerda con el coeficiente de correlación, y g) explique lo realizado a alguien que comprende la media, el desvío estándar, las puntuaciones Z y el coeficiente de correlación, pero que no sabe nada más sobre estadística.

Persona	**Destreza**	**Ansiedad**
1	1	10
2	1	8
3	2	4
4	4	−2
M	2	5
SD	1,22	4,58

4. Repita el ejercicio 3 resolviendo los puntos desde (a) hasta (f), pero prediciendo esta vez la destreza a partir de la ansiedad. Luego indique qué resultados son diferentes y cuáles son iguales a los obtenidos en el ejercicio 3 (**Nota**: SS_{Total} correspondiente a destreza es 6).

5. Ciertos psicólogos especializados en temas sociales que investigan temas relacionados con la justicia penal están interesados desde hace mucho tiempo en la influencia de varios factores en los sentimientos que despierta en el público el castigo impuesto a los criminales. Graham y sus colegas (1997) aprovecharon el muy famoso juicio de la estrella de fútbol americano O. J. Simpson para probar algunos temas básicos en este campo. Durante los primeros días después de que Simpson fue acusado de haber matado a su ex esposa, los investigadores formularon una serie de preguntas sobre el caso. Los investigadores estaban particularmente interesados en las respuestas de 177 individuos que creían probable que Simpson fuera culpable y, en especial, en la creencia que estas personas tenían respecto de la retribución: hasta qué punto estaban de acuerdo o no con la afirmación "el castigo debería hacer sufrir a Simpson lo que él hizo sufrir a otros". Los investigadores se centraron en una cantidad de posibles factores que influían sobre esas creencias. Los factores incluían el "control" (cuánto control creían ellos que Simpson tenía sobre sus acciones en el momento del crimen), la "responsabilidad" (cuán responsable por el crimen creían ellos que él era), cuánta "ira" sentían hacia él, cuánta "compasión" sentían por él, la "estabilidad" (hasta que punto creían que sus acciones representaban un comportamiento estable o temporario) y la "expectativa" (si creían que volvería a cometer un crimen de esa índole). El informe decía:

> La tabla [4-11] revela un apoyo parcial a nuestras hipótesis. Como era de esperarse, los predictores más importantes en cuanto al objetivo de retribución (hacer sufrir a Simpson) fueron las deducciones relacionadas con la responsabilidad y las emociones morales de ira y compasión. La estabilidad y la expectativa [...] fueron [predictores] relativamente débiles. (p. 337)

Explique estos resultados a una persona que comprende qué es una correlación pero que nunca ha escuchado hablar sobre análisis de regresión o regresión múltiple. (Haga referencia sólo a la parte de la tabla sobre retribución. La colum-

Persona	*A*	*B*	*C*	*D*	*E*	*F*	*G*	*H*	*I*	*J*
Control	1	0	0	0	0	0	1	0	1	−1
Responsabilidad	0	1	0	0	0	0	1	0	1	−1
Ira	0	0	1	0	0	0	1	1	1	−1
Compasión	0	0	0	1	0	0	0	1	1	−1
Estabilidad	0	0	0	0	1	0	0	0	1	−1
Expectativa	0	0	0	0	0	1	0	0	1	−1

na *t*, que se refiere a significación estadística de los resultados, se puede ignorar).

6. Sobre la base de la tabla 4-11 del ejercicio 5, escriba la ecuación de regresión para predecir la retribución. Luego determine la puntuación Z predicha para retribución correspondiente a las personas desde A hasta J, cuyas puntuaciones Z en cada variable predictora se detallan a continuación.

7. Pregunte a cinco alumnos de su mismo sexo (cada uno proveniente de una familia diferente) cuál es su altura y la de sus madres. Calcule el coeficiente de correlación, determine el modelo de predicción con puntuaciones originales para predecir la altura de una persona a partir de la altura de su madre y prepare un gráfico que muestre la recta de regresión. Finalmente, sobre la base del modelo de predicción determinado, prediga la altura de una persona de su mismo sexo cuya madre mide a) 5 pies, b) 5 pies y 6 pulgadas y c) 6 pies. (**Nota**: Convierta las pulgadas en decimales de los pies o resuelva todo el problema utilizando pulgadas).

Tabla 4-11.
Regresiones múltiples que predicen el castigo deseado a partir de variables de imputabilidad (estudio 1).

	Castigo deseado							
	Retribución		Rehabilitación		Protección		Disuasión	
Predictores	β	*t*	β	*t*	β	*t*	β	*t*
Control	–0,05	<1,07***	–0,05	<1,07***	–0,03	<1,07***	–0,15	1,90*
Responsabilidad	–0,17	–2,07***	–0,00	<1,07***	–0,04	<1,07***	–0,19	–2,15*
Ira	–0,30	–4,04***	–0,11	–1,54***	–0,03	<1,07***	–0,04	<1,00*
Compasión	–0,30	–3,68***	–0,39	–5,18***	–0,07	<1,07***	–0,13	–1,54*
Estabilidad	–0,01	<1,07***	–0,34	–4,85***	–0,19	2,33***	–0,04	<1,00*
Expectativa	–0,10	–1,33***	–0,06	<1,07***	–0,27	3,36***	–0,08	–1,04*
R^2		–0,27***		–0,37***		0,17***		–0,18*

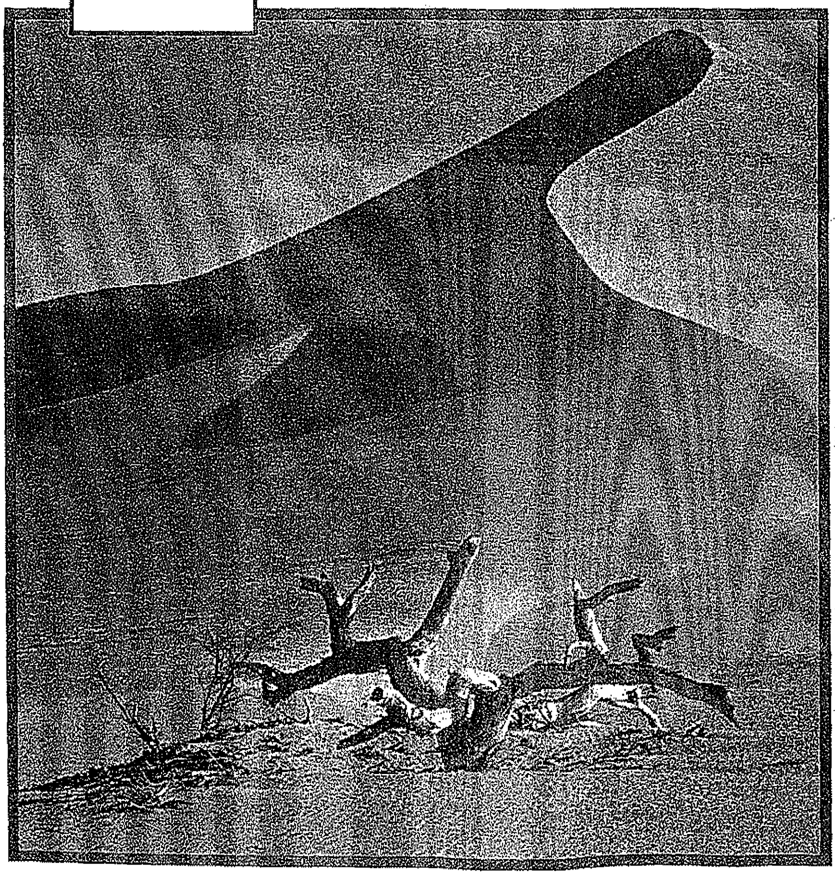
Nota: β = coeficiente de regresión estandarizado.

*$p < 0{,}05$; ***$p < 0{,}001$.

Fuente: Graham, S., Weiner, B., & Zucker, G. S. (*1997*), *tab. 4*. "Análisis del castigo deseado y reacción pública con respecto a O.J.Simpson basado en la imputabilidad". *Boletín de Psicología Social y de Personalidad [Personality and Social Psychology Bulletin]*, 23, 331–346. Copyright, 1997, por la Sociedad de Psicología Social y de Personalidad [Society for Personality and Social Psychology], Inc. Reimpreso con autorización de Sage Publications, Inc.

5 Algunos componentes clave de la estadística inductiva:

Curva normal, probabilidad y población versus muestra



Descripción del capítulo

- Distribución normal.
- Probabilidad.
- Muestra y población.
- Relación entre curva normal, probabilidad y muestra versus población.
- Controversias y limitaciones.
- Curvas normales, probabilidades, muestras y poblaciones según se describen en publicaciones científicas.
- Resumen.
- Términos clave.
- Ejercicios.
- Apéndice del capítulo: reglas de la probabilidad y probabilidades condicionales.

Comúnmente, los psicólogos realizan investigaciones para probar un principio teórico o la efectividad de algunos procedimientos prácticos. Por ejemplo, un psicofisiólogo podría medir los cambios en el ritmo cardíaco desde antes hasta después de resolver un problema difícil, y las mediciones podrían utilizarse luego para probar una teoría que predice que el ritmo cardíaco debería cambiar después de la solución exitosa de un problema. Un psicólogo especializado en temas sociales podría analizar la efectividad de un programa de reuniones vecinales con el fin de fomentar la conservación del agua. Tales estudios se realizan con un grupo determinado de personas que participan en la investigación, pero los investigadores utilizan la estadística inferencial para sacar conclusiones más generales sobre principios teóricos o procedimientos en estudio. Las conclusiones exceden el límite del grupo determinado de personas que participan en la investigación.

En este capítulo, al igual que en los capítulos 6, 7 y 8, presentamos la estadística inferencial, que establece los cimientos para la mayor parte de lo que resta del libro. El capítulo trata tres temas: curva normal, probabilidad y población versus muestra. Es un capítulo comparativamente corto, el cual prepara el camino para los próximos, que son más complejos.

DISTRIBUCIÓN NORMAL

En el capítulo 1 observamos que los gráficos de muchas distribuciones de variables estudiadas por los psicólogos (al igual que muchas otras distribuciones naturales) presentan forma de campana, aproximadamente simétrica y unimodal. Estos histogramas o polígonos de frecuencias con forma de campana se aproximan a una distribución matemática precisa e importante denominada

distribución normal o, simplemente, **curva normal.**[1] (Con frecuencia también se la denomina **distribución de Gauss**, en honor al astrónomo Karl Friedrich Gauss. Sin embargo, si su descubrimiento puede atribuirse a alguien, realmente debería atribuírsele a Abraham De Moivre, *véase* cuadro 5-1). La figura 5-1 muestra un ejemplo de curva normal.

¿Por qué la curva normal es tan común en la naturaleza?

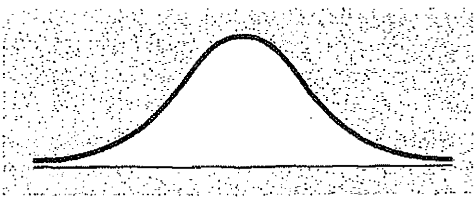
Tomemos, por ejemplo, la cantidad de letras elegidas al azar que determinada persona puede recordar con precisión en diversas pruebas (con diferentes letras elegidas al azar en cada oportunidad). En algunas pruebas, la cantidad de letras recordadas puede ser alta, en otras, baja, y en la mayoría, las cantidades serán intermedias. Es decir, es probable que la cantidad de letras elegidas al azar que una persona pueda recordar en diversas pruebas siga aproximadamente una curva normal. Supongamos que la persona tiene una capacidad básica para recordar de, digamos, siete letras, en este tipo de pruebas de memoria. Sin embargo, en alguna prueba en particular, el número real recordado se verá afectado por diversas circunstancias, tales como ruido en la habitación, estado de ánimo de la persona en ese momento, una combinación de letras confundidas inconscientemente con algún nombre familiar, una secuencia de letras elegidas al azar que resulta ser casi siempre la misma letra, etcétera.

Las distintas circunstancias se combinan y hacen que la persona recuerde más de siete palabras en algunas pruebas y menos de siete palabras en otras. No obstante, es muy probable que la combinación particular de estas circunstancias que ocurren en cualquier prueba se dé esencialmente por azar. Por lo tanto, en la mayoría de las pruebas las circunstancias positivas y negativas deberían prácticamente cancelarse unas a otras. Realmente no son muchas las chances de que ocurran todas las circunstancias negativas juntas en una prueba y ninguna circunstancia positiva.

Así, en general, la persona recuerda una cantidad media, una cantidad en la que todas las circunstancias contrapuestas se cancelan entre sí, y por eso son mucho menos comunes las cantidades muy altas o muy bajas de letras recordadas.

Esto crea una distribución que es unimodal, es decir, la mayoría de los casos están cerca del medio y los menos están en los extremos. También crea una distribución que es simétrica, porque cualquier valor puede estar tanto por arriba como por debajo del medio. Que la curva sea unimodal

Figura 5-1.
Una curva normal.



[1] La fórmula de la curva normal (cuando la media es 0 y el desvío estándar es 1) es:

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$$

donde $f(x)$ es la altura de la curva en el punto x, y π son las constantes matemáticas usuales (aproximadamente 3,14 y 2,72 respectivamente). Sin embargo, los psicólogos investigadores casi nunca utilizan esta fórmula, ya que está incluida en los distintos programas para computadoras que realizan cálculos estadísticos con curvas normales. Y cuando deben realizar el cálculo manualmente, cualquier información necesaria sobre la curva normal aparece en tablas en los libros de estadística (por ejemplo, la tabla B-1 en la última parte de este libro).

Cuadro 5-1.
De Moivre, el excéntrico desconocido que inventó la curva normal.

La curva normal es un tema central en estadística: es la base de la mayoría de las teorías y procedimientos estadísticos. Si existe una persona de la que pueda decirse que descubrió este principio fundamental de la materia, esa persona es Abraham De Moivre. Este era un protestante francés que llegó a Inglaterra a los 21 años de edad, huyendo de persecuciones religiosas en Francia, lugar que en 1685 se negaba a los protestantes todas sus libertades civiles. En Inglaterra, De Moivre entabló amistad con Isaac Newton, de quien se supone que muchas veces contestó preguntas diciendo: "Pregúntale a monsieur De Moivre, él sabe todo eso mejor que yo". Sin embargo, como era extranjero, De Moivre nunca pudo lograr la fama de los matemáticos británicos con los que trabajaba y que tanto lo respetaban.

De Moivre era principalmente experto en el azar. En 1733, escribió un "método de aproximación a la suma de los términos del desarrollo del binomio por una serie, de donde se deducen algunas reglas prácticas para estimar el grado de asentimiento que se debe otorgar a un determinado experimento". Su trabajo describía esencialmente la curva normal. Sin embargo, la descripción se realizaba sólo en forma de ley. En realidad, De Moivre nunca trazó la curva propiamente dicha, de hecho, no estaba muy interesado en ella.

Sus ideas sobre la distribución normal eran sólo una herramienta que desarrolló para calcular la probabilidad de que ocurra una cantidad determinada de veces un hecho para el que existen dos posibilidades, como por ejemplo, arrojar una moneda, cuando la cantidad de ensayos es muy grande. Hasta ese momento, se utilizaba una especie de cuadro denominado "triángulo aritmético" para aproximar el resultado. Pero el cuadro tenía que ser muy grande para, digamos, 1.000 tiros de moneda o 1.000 mediciones que podían o no ser exactas (estos inconvenientes fueron enfrentados en primer lugar por la astronomía y la psicofísica). De Moivre resolvió el problema con su "método de aproximación".

Con frecuencia se otorga el crédito por el descubrimiento de la curva normal a Pierre Laplace, un francés que permaneció en su lugar de origen, o a Karl Friedrich Gauss, alemán, o a Thomas Simpson, inglés. Todos ellos trabajaron en el problema de la distribución de errores en torno a una media, llegando incluso al punto de describir la curva o trazar aproximaciones de la misma. Pero aun sin dibujarla, De Moivre fue el primero en calcular las áreas debajo de la curva normal en 1, 2 y 3 desvíos estándar; Karl Pearson (de quien hablamos en el cuadro 14-1), un importante estadístico posterior, tuvo la fuerte convicción de que De Moivre era el verdadero descubridor de ese importante concepto.

En Inglaterra, De Moivre era muy apreciado tanto como hombre de letras como de números; conocía en profundidad a todos los clásicos y podía recitar escenas completas de su querido *Misántropo* de Moliere. Sin embargo, como reconocimiento a todo lo que sentía por su Francia natal, la Academia Francesa lo eligió, poco antes de su muerte, miembro **extranjero** de la Academia de Ciencias. No obstante, en Inglaterra no podía ser nombrado para una posición universitaria porque allí también era considerado extranjero. Su vida transcurrió en la pobreza, sin posibilidades siquiera de casarse. Durante los primeros años trabajó como profesor visitante de matemática. Más tarde, se hizo famoso por permanecer todo el día sentado en Slaughter's Coffee House* en Long Acre, esperando a los apostadores y aseguradores (dos profesiones igualmente inciertas y peligrosas antes de que se perfeccionaran las estadísticas), quienes le pagaban pequeñas sumas de dinero para que les calculara probabilidades.

La inusual muerte de De Moivre generó muchas leyendas. Trabajó mucho con series infinitas, que siempre convergen en cierto límite. Cierta historia relata que De Moivre comenzó a dormir quince minutos más cada noche hasta llegar a dormir constantemente, y después murió. Otra versión cuenta que su trabajo en el café lo llevó a tal grado de desesperación que simplemente se acostó a dormir hasta que murió. Es probable que la versión más exacta sea la que dice que sufría cierta enfermedad que lo hacía dormir cada vez más. En todo caso, a los ochenta años de edad sólo podía permanecer despierto cuatro horas por día, aunque se decía que en esas horas su tarea intelectual era más febril que nunca. Más tarde, sus horas de vigilia se redujeron a una hora por día, y luego desaparecieron. A los 87 años de edad, después de permanecer ocho días en cama, no despertó y fue declarado muerto por "somnolencia". Fue un hombre suficientemente inteligente, no sólo para inventar la curva normal sino también por irse de este mundo descansando.

Referencias: Pearson (1978); Tankard (1984).

y simétrica no garantiza que sea cercana a una curva normal; sus colas podrían ser demasiado altas o demasiado bajas. Sin embargo, puede demostrarse matemáticamente que, a la larga, si las circunstancias ocurren realmente al azar, el resultado será una perfecta curva normal. (La prueba puede encontrarse en algún texto de estadística matemática). Los estadísticos matemáticos llaman a este principio el **teorema del límite central**. Veremos más sobre este principio en el capítulo 7.

La curva normal y el porcentaje de casos ubicados entre la media y 1 y 2 desvíos estándar con respecto a ella

Debido a que la forma de la curva normal es estándar, existe un porcentaje conocido de valores por debajo y por encima en cualquier punto en particular. Por ejemplo, exactamente el 50% de los valores se encuentran por debajo de la media, porque en cualquier distribución simétrica la mitad de los valores se encuentran por debajo de la media. Más interesante es, como lo muestra la figura 5-2, que aproximadamente el 34% de los valores están siempre entre la media y 1 desvío estándar con respecto a ella. (A propósito, podemos observar que en la figura 5-2 el punto correspondiente a 1 desvío estándar en la curva normal coincide con el lugar de la curva en el que ésta comienza a ir más hacia fuera que hacia abajo).

Para ilustrar la utilidad del hecho de que la curva normal sea completamente estándar, consideremos los valores de CI. En muchas pruebas de inteligencia ampliamente utilizadas, el CI medio es 100, el desvío estándar es 16 y la distribución de valores de CI se considera aproximadamente normal (*véase* figura 5-3). Conocer la curva normal y el porcentaje de valores entre la media y 1

* N. de la Trad.: Un "*coffee house*" es similar a lo que nosotros llamamos "confitería" o "café".

desvío estándar por sobre la media nos permite saber que aproximadamente el 34% de las personas tienen registros CI entre 100 (la media de los CI) y 116 (el CI a 1 desvío estándar por encima de la media). Dado que la curva normal es simétrica, aproximadamente un 34% de las personas tienen un CI entre 100 y 84 (el valor ubicado a 1 desvío estándar por debajo de la media), y un 68% (34% + 34%) tiene un CI entre 84 y 116.

Observando la curva normal podemos observar algo más: existen muchos menos valores entre 1 y 2 desvíos estándar de la media que entre la media y 1 desvío estándar con respecto a ella. Aproximadamente el 14% de los valores se ubican entre 1 y 2 desvíos estándar por sobre la media (*véase* figura 5-2). De modo similar, siendo la curva normal simétrica, aproximadamente un 14% de los valores se encuentra entre 1 y 2 desvíos estándar debajo de la media. Por lo tanto, aproximadamente un 14% de personas tienen CI entre 116 (1 desvío estándar sobre la media) y 132 (dos desvíos estándar sobre la media).

Será muy útil recordar estos números: 34% y 14%. Las figuras indican el porcentaje de personas por encima y por debajo de cualquier valor en particular sólo con saber la cantidad de desvíos estándar por encima o por debajo de la media en que se encuentra dicho valor.

También es posible, a partir de un porcentaje, invertir el método y calcular la cantidad de desvíos estándar de la media a los que se encuentra determinada persona. Supongamos que nos informan que en determinada prueba una persona presentó un valor dentro del 2% más elevado. Suponiendo que los valores de la prueba tienen una distribución aproximadamente normal, la persona debe tener un valor al menos de dos desvíos estándar por encima de la media. Esto se debe a que del 50% de los valores ubicados por encima de la media, el 34% se encuentra entre la media y 1 desvío estándar por encima de ella; y otro 14% se encuentra entre 1 y 2 desvíos estándar sobre la media. Eso deja un 2% (es decir, 50% – 34% – 14% = 2%).

De manera similar, supongamos que estamos seleccionando animales para un estudio y necesitamos examinar su agudeza visual. Supongamos también que la agudeza visual está normal-

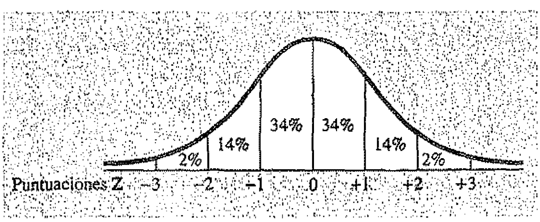


Figura 5-2. Curva normal con porcentaje aproximado de valores entre la media y 1, 2 y 3 desvíos estándar por encima y por debajo de la media..

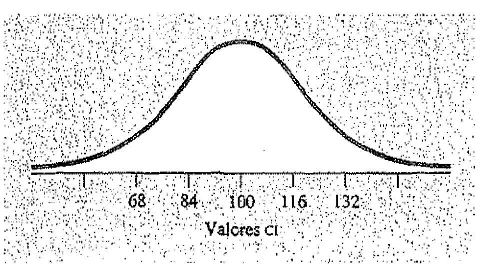


Figura 5-3. Distribución de valores CI de diversas pruebas estándar de inteligencia (con $M = 100$ y $SD = 16$).

mente distribuida, y que quisiéramos utilizar animales con un nivel de agudeza visual ubicada en los dos tercios centrales (un número cercano al 68%). En ese caso, seleccionaríamos animales que presentaron valores entre 1 desvío estándar por encima y 1 desvío estándar por debajo de la media. Si conociéramos la media y el desvío estándar de la prueba de agudeza visual, podríamos luego determinar las puntuaciones originales más bajas y más altas en cuanto a niveles de agudeza visual.

Tabla de la curva normal y puntuaciones Z

Los números 34% y 14% son referencias útiles para ser usadas como regla práctica de cálculo aproximado cuando un valor no se ubica exactamente a 1 ó 2 desvíos estándar de la media. Estos porcentajes nos dan una idea general del lugar en el que se ubica un valor en particular con respecto a los otros valores de la distribución. Sin embargo, en muchas situaciones aplicadas y de investigación, los psicólogos necesitan información más precisa. Afortunadamente, a causa de que la curva normal está definida de modo exacto, dicha precisión es posible. Es posible calcular, por ejemplo, el porcentaje exacto de valores entre dos puntos cualesquiera de la curva normal, no sólo aquellos que se encuentran exactamente a 1 ó 2 desvíos estándar de la media. Es decir, es posible determinar el porcentaje exacto de valores entre dos puntuaciones Z cualesquiera. Por ejemplo, exactamente un 68,59% de valores tienen puntuaciones Z entre +0,62 y –1,68; exactamente un 2,81% de registros tienen puntuaciones Z de entre +0,79 y +0,89, y así sucesivamente.

Estos porcentajes exactos pueden obtenerse, con la fórmula de la curva normal, aplicando el cálculo integral. Sin embargo, en la práctica, los psicólogos simplifican mucho el proceso. Los estadísticos han elaborado tablas para la curva normal que indican el porcentaje de valores entre la media (una puntuación Z igual a 0) y cualquier otra puntuación Z. Supongamos que necesitamos saber el porcentaje de valores entre la media y una puntuación Z de 0,62. Simplemente buscamos 0,62 en la tabla y ésta nos indica que el 23,24% de los valores se encuentran entre la media y esa puntuación Z.

En el apéndice B (tabla B-1) hemos incluido una **tabla de áreas bajo la curva normal**. Como se observa, la tabla consta de dos columnas. La primera incluye las puntuaciones Z, y la columna siguiente indica el porcentaje de valores entre la media y esa puntuación Z. Podemos observar también que las dos columnas se repiten varias veces en la página, por lo cual recomendamos ser cuidadosos al buscar los datos para no confundir las columnas. Además, la tabla indica sólo puntuaciones Z positivas, porque la curva normal es perfectamente simétrica y, por lo tanto, el porcentaje de valores entre la media y, digamos, una Z de +2,38, es exactamente igual al porcentaje de valores entre la media y un Z de –2,38.

En nuestro ejemplo, encontraríamos el número 0,62 en la columna correspondiente a "Z" y luego, justo al lado de ese número, en la columna correspondiente a "% entre media y Z", encontraríamos el número 23,24.

También podemos invertir el proceso y encontrar la puntuación Z que coincide con un porcentaje determinado de valores. Supongamos que nos informaran que el valor de Janice, en cuanto a creatividad, se encontraba dentro del 10% más elevado de los estudiantes de noveno grado. Demos por hecho, además, que los valores de creatividad siguen una curva normal. Podríamos calcular la puntuación Z de Janice de la siguiente manera: primero tendríamos que razonar que si ella se encuentra dentro del 10%, entonces el 40% de los estudiantes presentan valores entre el suyo y la media. (Existe un 50% por encima de la media y ella está entre el 10% más alto del total, es decir,que queda afuera un 40%). Luego, debemos mirar la columna de la tabla correspondiente a "% entre media y Z" hasta encontrar un porcentaje cercano al 40%. En este caso, el más cercano sería 39,97%. Finalmente, buscaríamos en la columna "Z" a la izquierda de este porcentaje. La puntuación Z correspondiente al 39,97% es 1,28. Conociendo la media y el desvío estándar de los valores de creatividad de estudiantes de noveno grado, podríamos calcular la puntuación original de Janice en la prueba. Lo haríamos transformando su puntuación Z de 1,28 en una puntuación original utilizando el método usual de conversión de puntuaciones Z en puntuaciones originales.

Procedimientos para calcular los porcentajes de valores a partir de puntuaciones originales y puntuaciones Z, utilizando la tabla de áreas bajo la curva normal

Basándonos en la explicación anterior, ahora podemos rever sistemáticamente los procedimientos para calcular los porcentajes de valores a partir de puntuaciones Z. Si estamos trabajando con puntuaciones originales, debemos convertirlos primero en puntuaciones Z utilizando los métodos descriptos en el capítulo 2, y luego proceder de la siguiente manera:

Primero realizamos un diagrama de la curva normal; marcamos el lugar en el que se ubica la puntuación Z y sombreamos el área con respecto a la cual estamos intentando encontrar el porcentaje. Luego estimamos el porcentaje del área sombreada sobre la base de la regla práctica, 50%–34%–14%. Al marcar la puntuación Z, debemos asegurarnos de ubicarla en el lugar correcto por encima o por debajo de la media, según sea positivo o negativo. Es importante hacer un diagrama del problema y realizar una estimación aproximada, ya que de ese modo es mucho menos probable que cometamos errores al realizar el cálculo más preciso.

Una vez que tenemos el diagrama y la estimación aproximada, podemos continuar con el proceso para encontrar el número exacto. El paso principal es buscar la puntuación Z en la columna "Z" de la tabla B-1 y buscar el porcentaje correspondiente en la columna "entre la media y Z" que se encuentra al lado. Si lo que buscamos es el porcentaje de valores entre la media y esa puntuación Z, esa sería nuestra respuesta final. Pero con frecuencia necesitaremos agregar un 50% a este porcentaje. Esto es necesario cuando la puntuación Z es positiva y buscamos el porcentaje total que se ubica por debajo de esa puntuación Z, o cuando la puntuación Z es negativa y buscamos el porcentaje total que se ubica por encima de esa puntuación Z. En otras ocasiones tendremos que restar al 50% el porcentaje obtenido. Esto es necesario cuando la puntuación Z es positiva y buscamos el porcentaje por encima de ella, o cuando la puntuación Z es negativa y buscamos el porcentaje por debajo de ella.

No es necesario memorizar las reglas que acabamos de mencionar. Es mucho más fácil realizar un diagrama del problema y calcular si el porcentaje que obtenemos de la tabla es correcto tal como está o si necesitamos sumar o restar un 50%.

Ejemplos

Analicemos algunos ejemplos utilizando valores de CI. Supongamos que una persona tiene un CI de 125. ¿Qué porcentaje de personas tiene mayores valores de CI? Antes de continuar necesitamos convertir la puntuación original en una puntuación Z. Suponiendo que la media es de 100 y el desvío estándar de 16, un valor CI de 125 es igual a una puntuación Z de +1,56. Ahora que tenemos la puntuación Z, el primer paso es realizar el diagrama. En la figura 5-4 hemos sombreado el área por encima de la puntuación Z de 1,56. Ahora queremos aproximar el porcentaje utilizando la regla 50%–34%–14%. Una puntuación Z de 1 tiene un 16% de valores por encima de ella (esto se debe a que hay un 34% de valores entre ella y la media, y existe un 50% de valores en total por encima de la media; es decir, que queda un 16% de valores por encima de 1 desvío estándar). Como vimos en uno de los ejemplos anteriores, por encima de una puntuación Z de 2 se ubica el 2% de los valores; por lo tanto, por encima de una puntuación Z de 1,56 habrá entre el 16% y el 2% de los valores.

Después de realizar el diagrama y estimar el porcentaje, estamos listos para calcularlo exactamente. En la tabla de áreas de la curva normal, 1,56 en la columna "Z" coincide con 44,06 en la columna "% entre la media y Z". Por lo tanto, el 44,06 % de las personas tiene valores de CI entre el CI medio y un CI de 125 (una puntuación Z de +1,56). En una curva normal, el 50% de las personas se encuentra por encima de la media. Dado que el 44,06% de las personas que se ubican por encima de la media se encuentran a su vez por debajo del CI de la persona analizada, queda un resto del 5,94% (50%–44,06%) de personas por encima del valor de la persona en cuestión. Esa es la respuesta a nuestro problema (representado por la figura 5-4). Cabe destacar que el porcentaje calculado se encuentra dentro del rango estimado utilizando la regla de aproximación del 50%–34%–14%.

Analicemos ahora a una persona con un CI de 95. ¿Qué porcentaje de personas presentan mayores valores de CI que la persona analizada? Siguiendo el procedimiento acostumbrado para convertir una puntuación original en una puntuación Z, un CI de 95 es igual a una puntuación Z de –0,31. La figura 5-5 muestra el diagrama para esta situación. Hemos sombreado el área de la curva superior a una puntuación Z de –0,31. La puntuación Z que analizamos se encuentra entre 0 y –1. Una puntuación Z igual a 0 tiene un 50% de los valores por encima de sí, y una puntuación Z de –1 tiene un 84% de los valores por encima de sí (esto se debe a que un 34% de los valores se ubican entre –1 y 0 y otro 50% se ubica por encima de 0, lo que sumado da un total de 84%). Por lo tanto, entre un 50% y un 84% de los valores se ubicarán por encima de la puntuación Z de –0,31.

Realicemos ahora el cálculo exacto. La tabla de áreas de la curva normal muestra que el 12,17% de los valores se encuentran entre la media y una puntuación Z de 0,31. Debido a que la

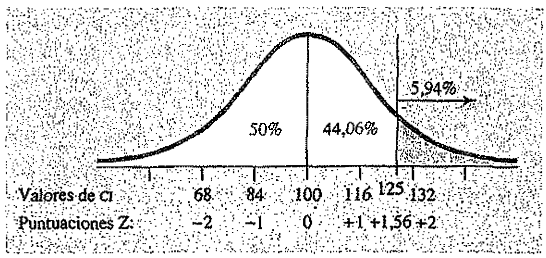


Figura 5-4. Distribución de valores de CI: la región sombreada corresponde al porcentaje de valores que se ubican por encima de un registro CI de 125.

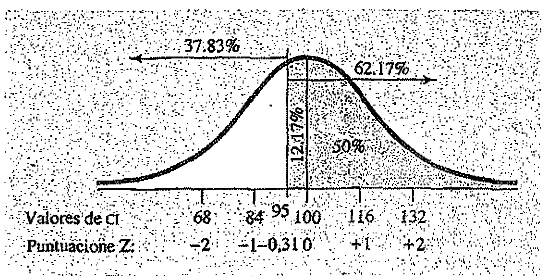


Figura 5-5. Distribución de valores de CI: la región sombreada corresponde al porcentaje de valores que se ubican por encima de un valor ci de 95.

curva normal es simétrica, ésta es también el área entre una puntuación Z de –0,31 y la media. Por lo tanto, el área total sobre –0,31 es 12,17% más el 50% que se ubica por encima de la media, lo que da un total de 62,17%. (El resultado se encuentra dentro de nuestro rango de aproximación del 50% al 84%).

Por otro lado, podemos observar también que el porcentaje de valores por **debajo** de una puntuación Z de –0,31 sería igual al 50% que se ubica por debajo de la media **menos** el 12,17% que se ubica entre la media y –0,31, dejando un total del 37,81% de los valores por debajo de la puntuación Z de –0,31.

Procedimientos para calcular puntuaciones originales y puntuaciones Z a partir de porcentajes de registros, utilizando la tabla de áreas bajo la curva normal

Obtener una puntuación Z a partir de un porcentaje es similar a la obtención de un porcentaje a partir de una puntuación Z. En ambos casos comenzamos realizando un diagrama del problema, sombreamos el porcentaje aproximado, y realizamos una estimación también aproximada de la puntuación Z utilizando los porcentajes del 50%–34%–14%. El resto del proceso es casi exactamente opuesto a ir de una puntuación Z a un porcentaje. Mirando el diagrama, calculamos el porcentaje entre la media y el lugar en el que comienza o termina el sombreado. Por ejemplo, si nuestro porcentaje es el 8% superior, entonces el porcentaje desde la media hasta donde comienza ese sombreado es igual al 42%. Si nuestro porcentaje es el 35% inferior, entonces el porcentaje desde la media hasta donde comienza el sombreado es del 15%. Si nuestro porcentaje es el 83% superior, entonces el porcentaje desde la media hasta donde termina el sombreado es del 33%.

Una vez que conocemos el porcentaje desde la media hasta donde comienza o termina el sombreado, buscamos el número más cercano que podamos encontrar en la columna de "% entre la media y Z" en la tabla de áreas bajo la curva normal. La puntuación Z en la columna "Z" al lado del porcentaje será nuestra respuesta, a menos que la puntuación Z que buscamos sea negativa. La mejor forma de saber si es positiva o negativa es a partir de la aproximación y del diagrama.

Si fuera necesario una respuesta final en puntuaciones originales, convertimos las puntuación Z en puntuaciones originales utilizando los métodos aprendidos en el capítulo 2.

Ejemplos

Una vez más, utilizaremos en nuestros ejemplos los valores de CI. ¿Qué CI necesitaría una persona para estar dentro del 5% superior? La figura 5-6 muestra nuestro diagrama, donde se observa que hemos sombreado el área que representa el 5% superior. Utilizando la regla del 50%–34%–14%, podemos adelantar que la puntuación Z correspondiente al 5% superior está entre +1 y +2. El cálculo que realizamos fue el siguiente: del 50% que se encuentra por encima de la media, el 34% se ubica entre la media y 1 desvío estándar, con lo cual queda un 16% superior a 1 desvío estándar. Sin embargo, dado que hay un 14% entre 1 y 2 desvíos estándar, queda sólo un 2% superior a 2 desvíos estándar.

Con respecto a la puntuación Z exacta, primero averiguamos el porcentaje entre la media y el lugar en el que empieza nuestra área sombreada. En este caso, si el 50% de las personas tienen valores de CI superiores a la media, al menos un 45% de las personas presentan valores de CI ubicados entre la persona en cuestión y la media (50% – 5% = 45%). Buscando en la columna "% entre la media y Z" en la tabla de áreas bajo la curva normal, el valor más cercano al 45% es 44,95% (también podríamos utilizar el 45,05%). Este porcentaje coincide con una puntuación Z de 1,64 en la columna "Z". Tal como lo esperábamos según nuestra aproximación inicial, la respuesta se ubica entre +1 y +2.

Para averiguar la puntuación original podemos utilizar la fórmula del capítulo 2: $X = M + (Z)(SD)$. Con un CI medio de 100 y un desvío estándar de 16, llegaríamos a la conclusión de que para estar dentro del 5% superior, una persona necesitaría un CI de por lo menos 126,24.

Analicemos ahora qué valor de CI estaría dentro del 2,5% inferior. La figura 5-7 representa nuestro diagrama del problema, sombreado en la parte correspondiente al 2,5% inferior. El 2% inferior de una curva normal comienza en el segundo desvío estándar inferior a la media (igual que el 2% superior comienza en +2). Por lo tanto, podemos estimar que nuestra respuesta estará en algún punto cercano al −2. En términos más precisos, el 2,5% inferior significa que, al menos, el 47,5% de las personas presentan valores de CI ubicados entre el valor de CI que pretendemos determinar y la media (50% − 2,5% = 47,5%). En la tabla de áreas de la curva normal, el 47,5% en la columna "% entre la media y Z" coincide con una puntuación Z de 1,96. Debido a que estamos buscando una puntuación Z por debajo de la media, el número ubicado en la tabla se transforma en −1,96 (un número bastante cercano a nuestra estimación de −2). Al convertir este resultado en una puntuación original, el CI correspondiente al 2,5% inferior resulta ser un CI de 68,64.

PROBABILIDAD

El objetivo de la mayor parte de las investigaciones psicológicas es probar la veracidad de una teoría o la efectividad de un procedimiento. Pero la investigación científica de cualquier tipo sólo puede llegar a la conclusión de que la veracidad o efectividad resultan más o menos probables; no puede proporcionarnos el lujo de la certeza. La probabilidad es muy importante para las ciencias. En particular, es muy importante para la estadística inferencial, es decir, para los métodos utilizados por los psicólogos para sacar conclusiones sobre teorías o procedimientos aplicados a partir de los resultados obtenidos en investigaciones.

La probabilidad ha sido estudiada durante siglos por matemáticos y filósofos y, sin embargo, aún en nuestros días el tema despierta todo tipo de controversias. Afortunadamente, sólo necesita-

Figura 5-6. Puntuación Z y puntuación original CI correspondientes al 5% superior.

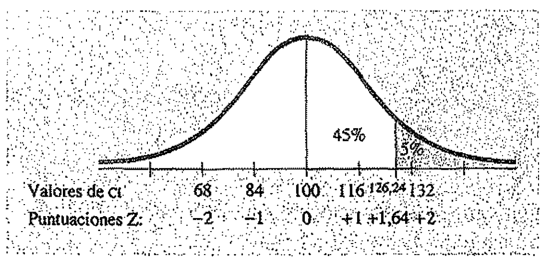
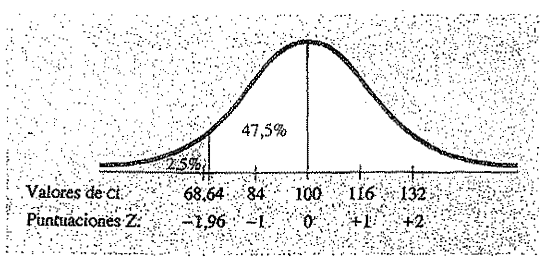


Figura 5-7. Puntuación Z y puntuaciones originales CI correspondientes al 2,5% inferior.



mos conocer unas pocas ideas clave para comprender y realizar los procedimientos de inferencia estadística que aprenderemos en este libro. Esos pocos puntos clave no son muy complejos; de hecho, algunos alumnos los consideran intuitivamente obvios.

Interpretaciones de la probabilidad

En estadística, generalmente definimos **probabilidad** como "la frecuencia relativa con que esperamos que suceda un determinado resultado". Un **resultado** es la consecuencia de un experimento (o de casi cualquier situación en la que la consecuencia no se conoce de antemano, como puede ser que una moneda caiga cara arriba o que llueva mañana). La **frecuencia** indica cuántas veces sucede determinado hecho. La **frecuencia relativa** es la cantidad de veces que determinado hecho sucede en relación con la cantidad de veces que podría haber sucedido, es decir, la razón entre la cantidad de veces en que algo sucede y la cantidad de veces que podría haber sucedido. (Una moneda podría caer cara arriba 8 veces en 12 tiros, con una frecuencia relativa de 8/12 ó 2/3). La **frecuencia relativa esperada** indica lo que esperaríamos que suceda a largo plazo si repitiéramos el experimento muchas veces. (En el caso de una moneda, esperaríamos que en el largo plazo la moneda caiga cara hacia arriba una de cada dos veces). A esto se lo denomina **interpretación de la probabilidad como la frecuencia relativa a largo plazo.**

También utilizamos la probabilidad para transmitir en qué medida estamos seguros de que sucederá un hecho en particular. A esto se lo denomina **interpretación subjetiva de probabilidad.** Supongamos que decimos que existe un 95% de probabilidad de que nuestro restaurante favorito esté abierto esta noche. Podríamos estar aplicando una especie de interpretación de frecuencia relativa, lo cual implicaría que si verificáramos si ese restaurante estuvo abierto muchas veces en días como hoy, descubriríamos que en un 95% de esos días efectivamente estuvo abierto. Sin embargo, lo que en realidad queremos decir es probablemente más subjetivo: en una escala del 0% al 100%, calificaríamos nuestra confianza en que el restaurante estará abierto con un 95%. Para decirlo de otro modo, sentiríamos que una apuesta sería justa si se basara en que las chances de que el restaurante va a estar abierto son del 95%.

La interpretación que uno adopte no afecta la forma de calcular las probabilidades. Presentamos estos conceptos aquí por dos razones. Primero, queríamos dar una idea un poco más profunda del significado del término **probabilidad**, el cual ocupará un lugar destacado durante el resto del aprendizaje de estadística, aun cuando, como sucede a menudo, este conocimiento más profundo no se convierta en un dogma. En segundo lugar, es de crucial importancia familiarizarse con ambas interpretaciones para comprender algunas de las controversias más encendidas dentro de la estadística, una de las cuales presentaremos al final de este capítulo.

Cálculo de probabilidades

En las aplicaciones estadísticas, las probabilidades se calculan con una proporción de resultados exitosos, es decir, la cantidad de resultados favorables dividida por la cantidad de resultados posibles.

Analicemos la probabilidad de que al lanzar una moneda ésta caiga cara hacia arriba. De los dos resultados posibles (obtener cara o cruz), existe un resultado favorable (obtener cara), es decir, una probabilidad de 1/2 ó 0,5. Si tiramos un sólo dado, la probabilidad de sacar un 2 (o cualquier otra cara del dado) es de 1/6 ó 0,17. Es decir, de los seis resultados posibles hay sólo un resultado favorable en particular. La probabilidad de tirar un dado y obtener un número 3 o menor es de 3/6, ó 0,5. De seis resultados posibles existen tres resultados favorables (un 1, un 2 o un 3).

Analicemos un ejemplo un poco más complicado. Supongamos que en una clase hay 200 personas, y que 30 son estudiantes avanzados. Si eligiéramos alguien de la clase al azar, la probabilidad de escoger un estudiante avanzado sería *30/200*, ó *0,15*. Es decir, de 200 resultados posibles existen 30 resultados favorables (elegir un estudiante avanzado).

Rango de probabilidades

Las probabilidades son razones (la cantidad de resultados favorables sobre el total de resultados posibles). *Esta razón no puede ser menor que 0 ni mayor que 1. Expresada en porcentajes*, va del 0% al 100%. Algo que no tiene chances de ocurrir tiene probabilidad 0 y algo que ocurrirá con certeza tiene probabilidad 1. Cuando un hecho no puede ocurrir o es **imposible**, tiene probabilidad 0, pero cuando la probabilidad de un hecho es baja, digamos un *5%* o incluso un *1%*, el hecho es **improbable o poco probable**, pero no imposible.

Las probabilidades expresadas con símbolos

La probabilidad se simboliza generalmente con la letra p. El número real que representa una probabilidad por lo general es un decimal, aunque a veces se utilizan fracciones o porcentajes. Así, si las chances son 50–50 usualmente se expresa por escrito $p = 0{,}5$, aunque también podría expresarse $p = 1/2$ ó $p = 50\%$. También es común ver la probabilidad expresada por escrito como "menos que" algún número, utilizando el signo "menos que" (<). Por ejemplo, "$p<0{,}05$" significa "la probabilidad es menor a un 5%".

Reglas de probabilidad

Como mencionamos anteriormente, nuestra exposición sólo trata el tema de la probabilidad de modo superficial. Uno de los aspectos que no hemos tenido en cuenta son las reglas para calcular probabilidades relacionadas con experimentos o resultados múltiples (por ejemplo, ¿cuál es la probabilidad de tirar una moneda dos veces y que en ambos casos caiga cara hacia arriba?). Estas normas se denominan reglas de probabilidad y representan un papel muy importante en las bases matemáticas de muchos aspectos de la estadística. Sin embargo, no es necesario conocer las reglas de probabilidad para comprender el material cubierto por este libro. Más aún, estas reglas rara vez se utilizan directamente al analizar los resultados de investigaciones psicológicas. De todos modos, en algunas publicaciones científicas podríamos ocasionalmente encontrarnos con referencias a esos procedimientos; por lo tanto, en el apéndice del capítulo describimos las dos reglas de probabilidad más comúnmente mencionadas.

Probabilidad y la distribución normal

Hasta ahora hemos tratado principalmente probabilidades de hechos puntuales que podrían suceder como no suceder. También podemos hablar de hechos más generales que podrían o no suceder, como por ejemplo, lanzar un dado y obtener un 3 ó un número menor. Otro ejemplo sería la probabilidad de elegir a alguien que tenga entre 30 y 40 años de edad, en una calle de la ciudad.

Si pensamos en la probabilidad en términos de cantidad de resueltados favorables, sobre la cantidad de casos posibles, la probabilidad puede equipararse adecuadamente con las distribuciones de frecuencias (*véase* capítulo 1). Analicemos la distribución de frecuencias que aparece en el histograma de la figura 5-8. Del total de 50 números, 10 son 7 ó mayores. Si estuviéramos selec-

Cuadro 5-2.
Pascal comienza a desarrollar la teoría de la probabilidad en las mesas de juego y más tarde aprende a apostar a Dios.

Mientras que en Inglaterra se utilizaba la estadística para tener una idea del índice de mortalidad y para probar la existencia de Dios (*véase* cuadro 1-1), los franceses y los italianos desarrollaron la estadística alrededor de las mesas de juego. Existía un problema en particular, denominado "problema de los puntos" (el reparto de las apuestas en un juego después de interrumpirlo). Si se pensaba realizar cierta cantidad de juegos, ¿qué parte de las apuestas debería llevarse cada jugador según el porcentaje ya jugado?

El problema fue tratado, al menos, a partir del año 1494 por Luca Pacioli, un amigo de Leonardo da Vinci. Pero permaneció sin resolverse hasta el año 1654, cuando fue presentado a Blaise Pascal por el Caballero de Méré. Pascal, un niño prodigio francés, asistía a reuniones de los más famosos matemáticos franceses adultos, y a los 15 años de edad probó un importante teorema de geometría. Junto con Pierre Fermat, otro famoso matemático francés, Pascal resolvió el problema de los puntos, y al hacerlo, comenzó con el desarrollo de la teoría de la probabilidad y con el trabajo que terminaría dando como resultado la curva normal.

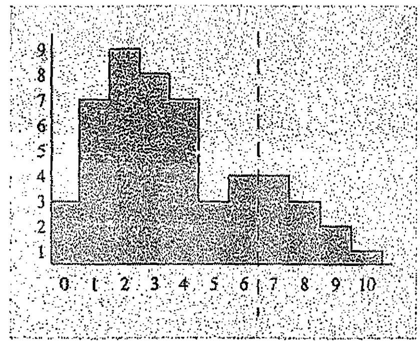
No mucho tiempo después de resolver este problema, Pascal se convirtió repentinamente en un religioso tan devoto como los estadísticos ingleses. El coche en el que viajaba cruzó desbocado un puente y Pascal se salvó de ahogarse sólo porque en el último instante se rompieron los tirantes del atelaje. Consideró que esto era una advertencia para que abandonara su trabajo matemático y se dedicara a la escritura religiosa, formulando más tarde la "Apuesta de Pascal": el valor de un juego está dado por el valor del premio multiplicado por las probabilidades de ganarlo; por lo tanto, aun cuando las probabilidades de que exista Dios sean bajas, deberíamos apostar por la afirmación de su existencia, ya que el valor del premio es infinito, mientras que el valor de no creer es sólo placer mundano finito.

Referencia: Tankard (1984).

cionando al azar personas, cada una con un número de estos asignados, habría 10 posibilidades (resultados favorables) de 50 (todos los resultados posibles) de seleccionar una que tuviera asignado un número igual a 7 ó mayor. Por lo tanto, $p = 10/50 = 0{,}2$

La distribución normal también puede considerarse como una distribución de probabilidades. La curva normal representa a una distribución de frecuencias en la que se conoce la proporción de valores entre dos puntuaciones Z cualesquiera. Como hemos visto, la proporción de valores entre dos puntuaciones Z cualesquiera es la mismo que la probabilidad de seleccionar un valor entre esas dos puntuaciones Z. Por ejemplo, la probabilidad de que un valor se encuentre entre la media y una puntuación de +1 (1 desvío estándar por encima de la media) es de aproximadamente un 34%, es decir, $p = 0{,}34$.

Figura 5-8. Distribución de frecuencias (en forma de histograma) de la selección de 50 números, en las que la probabilidad de elegir el 7 ó un número mayor es $p = 0,2$ (10/50).



Es probable que lo que estamos diciendo haya sido obvio desde el principio. En algún sentido, el hecho de que la curva normal pueda representar tanto a una distribución de frecuencias como a una distribución de probabilidades, es meramente un tema técnico. Sólo lo mencionamos para que no haya confusiones más adelante, cuando hagamos referencia a la probabilidad de que un valor esté en un intervalo, como el área sobre él bajo la curva normal.

MUESTRA Y POBLACIÓN

Presentaremos algunas ideas importantes utilizando el ejemplo de las habas. Supongamos que estamos cocinando una olla con habas y probamos una cucharada para ver si están listas. En este ejemplo, la olla con habas es la **población**, la cantidad completa de elementos que nos interesan. La cucharada es la **muestra**, la parte de la población sobre la cual realmente tenemos información. La figura 5-9 grafica el ejemplo.

En la investigación psicológica, generalmente estudiamos muestras, no de habas sino de individuos. Una muestra podría consistir en 50 mujeres canadienses que participan en determinado experimento; la población que uno podría tener el propósito de reflejar serían todas las mujeres canadienses. En un sondeo de opinión, podríamos seleccionar 1.000 personas de toda la población con edad para votar de un determinado país y preguntarles por quién votarían. Las opiniones de esas 1.000 personas conforman la muestra. Las opiniones del todo el público votante en ese país, respecto de quienes los encuestadores van a generalizar sus resultados, son la población (*véase* figura 5-10)[2].

¿Por qué se analizan muestras y no poblaciones?

Como hemos visto, los investigadores realizan estudios para averiguar algo sobre una determinada población. Por lo tanto, sus resultados serían mucho más precisos si pudieran estudiar la población completa, en lugar de un subgrupo de esa población. Sin embargo, en la mayoría de las

[2] Estrictamente hablando, los términos **población** y **muestra** se refieren a una serie de valores (números o mediciones), no a los participantes de la investigación que fueron medidos. Por lo tanto, en el primer ejemplo, la muestra está formada en realidad por los valores observados de las 50 mujeres canadienses, y no por las 50 mujeres, mientras que la población está conformada por los valores que se obtendrían si se midieran todas las mujeres canadienses.

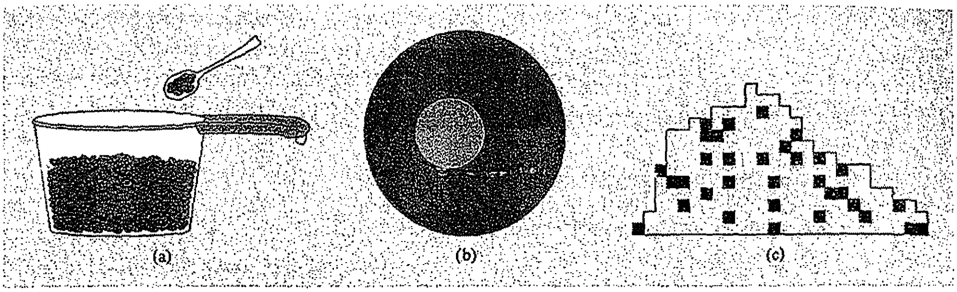


Figura 5-9. Poblaciones y muestras: en (a), toda la olla con habas es la población, y la cucharada es la muestra. En (b), todo el círculo mayor es la población y el círculo que se encuentra dentro de éste es la muestra. En (c), el histograma se refiere a la población, y los valores sombreados tomados en conjunto forman la muestra.

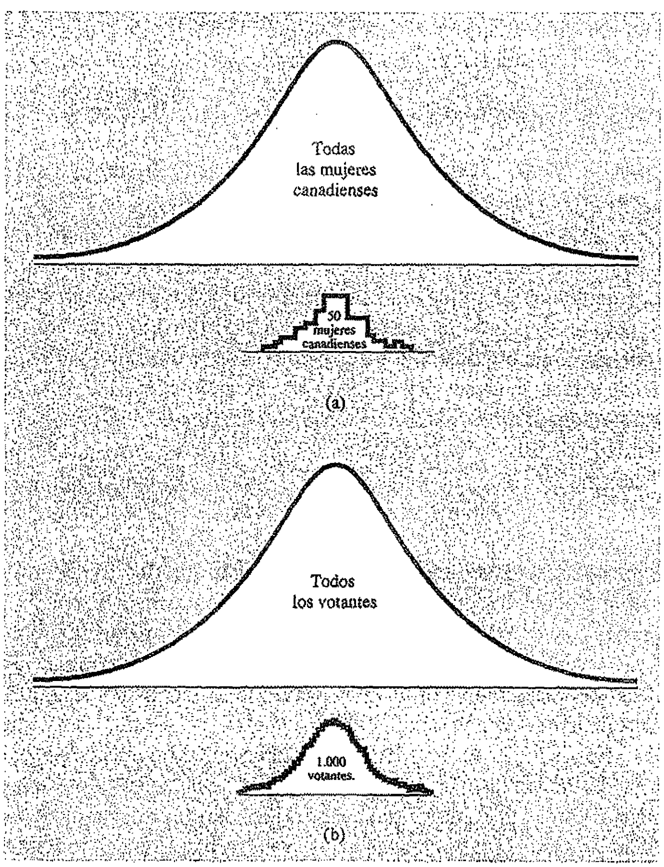


Figura 5-10. Otros ejemplos de poblaciones y muestras. En (a), la población está conformada por los valores de todas las mujeres canadienses, y una muestra está formada por los 50 valores observados particulares de las 50 mujeres canadienses estudiadas. En (b), la población está conformada por las preferencias en cuanto a voto de toda la población en edad de votar de un país, y una muestra está formada por las preferencias en cuanto a voto de las 1.000 personas en edad de votar de ese país, a las que se les realizó la encuesta.

investigaciones esto no es practicable. Aun más importante, el sentido de la investigación es, por lo general, poder realizar generalizaciones o predicciones acerca de hechos que están más allá de nuestro alcance. No sería investigación científica probar nuestros tres automóviles para ver cuál tiene mayor rendimiento por milla, a menos que nuestro objetivo fuera probar algo acerca del rendimiento por milla en cuanto a esos modelos de automóviles en general. En otras palabras, un investigador podría realizar un experimento sobre la manera en que las personas almacenan palabras en la memoria a corto plazo, utilizando 20 alumnos como participantes en el experimento. Pero el objetivo del experimento no es averiguar cómo responden esos 20 alumnos en particular a las condiciones experimentales. Más bien, el objetivo es aprender algo acerca de la memoria humana en esas condiciones.

La estrategia de la mayoría de las investigaciones psicológicas es estudiar una muestra de individuos considerados representativos de la población general (o de alguna población determinada en la que estamos interesados). De manera más realista, los investigadores intentan estudiar a aquellas personas que no difieren de la población general de ningún modo sistemático que pudiera influir en el tema en análisis.

Por lo tanto, en la investigación psicológica (y en casi todas las investigaciones científicas), lo que se analiza es la *muestra*. La población es algo desconocido sobre lo cual los investigadores sacan conclusiones sobre la base de la muestra. La mayor parte de lo que aprenderemos en lo que resta del libro está basado en la importante tarea de sacar conclusiones acerca de poblaciones, tomando como referencia la información obtenida a partir de las muestras.

Métodos de muestreo

Dado que existen tantas formas de seleccionar una muestra para un proyecto de investigación en particular, en el apéndice A presentamos una exposición sobre varios de estos métodos (*véase* también cuadro 5-3). Brevemente, podemos decir que en la mayoría de los casos el método ideal para seleccionar una muestra de estudio se denomina **selección aleatoria**. El investigador consigue una lista completa de los miembros de la población y selecciona al azar una cantidad para analizar. Un ejemplo del método de selección aleatoria sería escribir cada nombre en una pelotita de ping pong, colocar las pelotitas en un gran recipiente, sacudirlo y vendarle los ojos a una persona para que seleccione la cantidad necesaria. (En la práctica, la mayoría de los investigadores utilizan una lista de números aleatorios generada por computadora. La manera en que las computadoras o las personas pueden crear una lista de números realmente aleatorios es una cuestión interesante en sí misma que analizaremos en el cuadro 15-1).

Es importante distinguir la selección verdaderamente aleatoria de lo que podríamos denominar **selección casual**, como por ejemplo, elegir a quien esté disponible o primero en la lista. Utilizando el método de selección casual, es sorprendentemente fácil elegir accidentalmente un grupo de personas para estudiar que sean en realidad muy diferentes a la población en su conjunto. Analicemos el caso de un estudio de actitud para con un profesor de estadística. Supongamos que recolectamos la información para análisis de entre aquellos que en clase se sientan cerca de determinado alumno. Ese análisis estaría afectado por todos los factores que influyen en la elección del asiento, algunos de los cuales tienen que ver precisamente con el tema que estamos analizando, como por ejemplo, en qué medida los alumnos están conformes con el profesor o con la clase. (De modo similar, pedirle información a las personas que se sientan cerca de determinado alumno, daría como resultado obtener opiniones más similares a las de ese alumno, de lo que resultarían las opiniones obtenidas por medio de una verdadera muestra aleatoria).

Desafortunadamente, en la investigación psicológica sólo es posible estudiar muestras verdaderamente aleatorias en algunas ocasiones. La mayor parte del tiempo, de hecho, se realizan análisis con aquellos que quieren o pueden participar de una investigación. En el mejor de los casos, como ya observamos, el investigador intenta analizar una muestra de individuos de quienes no se conozca ningún dato que pueda hacerlos sistemáticamente no representativos de la población que se intenta analizar. Por ejemplo, supongamos que se realiza un estudio acerca de un proceso que puede dar diferentes resultados según las distintas edades de las personas. En ese caso, el investigador puede intentar incluir en el análisis personas de todas las edades. Otra alternativa es que el investigador sea cuidadoso al sacar las conclusiones, para que estas se refieran sólo al grupo correspondiente a la edad estudiada.

Terminología estadística relacionada con muestras y poblaciones

La media, la varianza y el desvío estándar de una población se denominan **parámetros poblacionales**. Generalmente se desconocen los parámetros de una población, y sólo pueden estimarse a partir de lo que sabemos acerca de una muestra tomada de esa población. No probamos todas las habas, sino sólo una cucharada. "Están listas" es una estimación referida a toda la olla.

Para recordar esta diferencia, resulta útil saber que los parámetros poblacionales generalmente se simbolizan con letras griegas. El símbolo que representa la media de una población es μ, la letra griega "mu"; el símbolo que representa la varianza de una población es σ^2, y el símbolo que representa su desvío estándar es σ, la letra griega "sigma" minúscula. Estos signos no aparecerán con mucha frecuencia, excepto mientras estudiamos estadística, ya que, como dijimos anteriormente, los investigadores rara vez conocen los parámetros poblacionales.

La media, la varianza y el desvío estándar que calculamos según los registros de una muestra se denominan **estadísticos muestrales**. Un estadístico muestral se calcula a partir de información conocida. Los estadísticos muestrales son los que hemos estado calculando hasta ahora y se representan por los símbolos que hemos estado utilizando: M, SD^2 y SD. La tabla 5-1 resume los diversos símbolos.

Tabla 5-1.
Parámetros poblacionales y estadísticos muestrales.

	Parámetro poblacional (usualmente desconocido)	**Estadístico muestral** (Calculado a partir de datos conocidos)
Base:	Valores de la población completa	Valores sólo de la muestra
Símbolos:		
Media	μ	M
Desvío estándar	σ	SD
Varianza	σ^2	SD^2

Cuadro 5-3. Sondeos, encuestas y la costosa "muestra gratis" de 1948.

Como lectores de resultados de sondeos o encuestas que difunden los medios, llegó el momento de estar mejor informados. En líneas generales, los resultados de encuestas públicas realizadas correctamente se presentan junto con alguna afirmación tal como: "Fuente: encuesta telefónica a 1.000 adultos norteamericanos realizada el 4 y 5 de junio. Error de muestreo ±3%", escrita en letra menuda. ¿Qué significa todo esto?

La encuesta Gallup es un muy buen ejemplo (Gallup, 1972), y no existe mejor momento para comenzar que en 1948, cuando las tres mejores empresas encuestadoras, Gallup, Crossley (para los periódicos Hearst), y Roper (para Fortune), predijeron equivocadamente la victoria de Thomas Dewey contra Harry Truman para la presidencia de los EE.UU. Sin embargo, la predicción de Gallup se basó en 50.000 entrevistas, y la de Roper en 15.000. Por el contrario, para predecir la victoria de George Bush en las elecciones presidenciales de EE.UU. de 1988, Gallup utilizó sólo 4.089 entrevistas. Desde 1952, los encuestadores no han utilizado nunca más de 8.144 entrevistas, pero con muy bajo error y sin equivocaciones garrafales. ¿Qué ha cambiado?

El método utilizado antes de 1948, y nunca repetido desde entonces, se denominaba "muestreo por cuotas". A los entrevistadores se les asignaba una cantidad fija de personas a entrevistar, con cupos estrictos para completar en todas las categorías consideradas importantes, tales como lugar de residencia, sexo, edad, raza y nivel económico. Sin embargo, si cumplían con estos datos específicos, podían entrevistar a quien quisieran. En los Estados Unidos, los republicanos generalmente solían ser más fáciles de entrevistar, ya que era más probable que tuvieran teléfono y residencias permanentes, y que sus viviendas fueran de mejor nivel y que estuvieran ubicadas en mejores vecindarios. Antes de 1948 no se le había dado importancia a este leve sesgo. Los demócratas habían estado ganando durante años por grandes márgenes. En 1948, la elección fue muy reñida, y el sesgo republicano produjo el vergonzoso error que cambió para siempre los métodos de sondeo de datos.

Desde 1948, Gallup y las otras empresas encuestadoras han utilizado lo que se denomina "método probabilístico". El muestreo al azar simple es la forma más pura del método probabilístico, pero aplicado a una encuesta acerca de las elecciones presidenciales de los EE.UU. requeriría elegir los nombres de entre una lista de todos los votantes de la nación, lo cual significaría una cantidad de personas demasiado grande. Luego debería encontrarse a cada persona seleccionada en diferentes y alejados lugares. Por lo tanto, se utiliza el "muestreo de agrupación de escenarios múltiples". Para describirlo someramente, se dividió a los Estados Unidos en siete estratos del tamaño de una comunidad, desde las grandes ciudades hasta las zonas rurales; estos grupos se dividieron en siete regiones geográficas (Nueva Inglaterra, Atlántica Media, etc.), que a su vez se dividieron en zonas menores de igual tamaño, y luego se tomaron manzanas de cada una de las zonas; siendo las probabilidades de selección proporcionales a la cantidad de población o viviendas. Finalmente, se dio al entrevistador un punto de partida elegido al azar en el mapa y se le pidió que siguiera determinada dirección, que pasara por cada casa siguiendo el orden en que estaban ubicadas y que preguntara por el hombre más joven mayor de 18 años, o si no había ningún hombre en la casa, por la mujer de mayor edad que tuviera más de 18. (Se ha descubierto que esta es la mejor manera de compensar la tendencia de que los hombres jóvenes, todos los hombres y por último las mujeres mayores, en ese orden, no se encuentren en sus casas y, por lo tanto, no sean representados adecuadamente).

En realidad, dado que las encuestas telefónicas cuestan aproximadamente la tercera parte de lo que cuestan las encuestas puerta a puerta, y que la mayoría de la gente en la actualidad tiene teléfono, con lo que resulta reducido el sesgo de este método en favor de la gente adinerada como sucedía en la época de Truman, los llamados telefónicos son actualmente el método preferido para realizar sondeos. Los llamados telefónicos también permiten que las computadoras disquen al azar a través de un complicado sistema denominado RDD (*Random Digit Dialing*, Discado de dígitos aleatorio) que, a diferencia de los directorios telefónicos, incluye números no inscriptos.

Ya sea que la encuesta se realice por teléfono o cara a cara, habrá un 35% de las personas que no responderán a pesar de los tres intentos de contactarlas. Esto crea otro sesgo más, que se toma en cuenta a través de preguntas acerca de la cantidad de tiempo que la perona pasa en su casa, para dar así un poco más de importancia relativa a las respuestas de aquellos que pudieron ser contactados pero que, por lo general, pasan menos tiempo en su casa, y compensar a aquellos que no pudieron ser contactados.

Ahora sabemos bastante sobre sondeos de opinión. Aunque hemos dejado sin respuesta dos importantes preguntas: ¿Por qué se incluyen sólo 1.000 entrevistas en un sondeo que pretende representar a todos los adultos de EE.UU.?, y ¿qué significa el término error de muestreo? Para contestar estas preguntas debemos esperar hasta el capítulo 7 (cuadro 7-1).

RELACIÓN ENTRE CURVA NORMAL, PROBABILIDAD Y MUESTRA VERSUS POBLACIÓN

Como dijimos anteriormente, en la mayoría de las investigaciones no conocemos los parámetros poblacionales. Sin embargo, generalmente suponemos que la población es aproximadamente normal. Por lo tanto, los investigadores recolectan usualmente información acerca de una muestra para realizar inferencias probabilístcas acerca de los parámetros de una población normalmente distribuida.

Analicemos un experimento realizado para averiguar si los alumnos aprenden más cuando estudian todo de una vez o cuando el estudio se reparte a lo largo de un periodo de tiempo. Se seleccionaron 60 alumnos al azar para participar en la investigación. A una mitad, escogida también al azar, se le asigna la tarea de estudiar todo de una vez y, a la otra mitad, la de estudiar la misma cantidad de horas repartidas a lo largo de varias semanas. Al final de esas semanas, se toma un examen a ambos grupos. El resultado es que existe una diferencia entre los dos grupos en cuanto a las calificaciones medias en la prueba.

Ahora bien, analicemos el experimento en función del lenguaje que hemos estado utilizando en este capítulo. El grupo que estudió todo de una vez es una muestra. Esta muestra tiene el propósito de representar el desempeño de los alumnos en general, si tuvieran que estudiar todo de una sola vez. Es decir, esta muestra representa a una población hipotética de alumnos a los que se les asigna el estudio de un tema todo de una sola vez. El grupo que estudió durante un periodo de *tiempo* es otra muestra. Esta muestra pretende representar el desempeño de los alumnos en general a quienes se les asignó estudiar a lo largo de un periodo de tiempo. Por lo tanto, esta muestra representa una población hipotética de alumnos a quienes se les asigna estudiar un tema a lo largo de un periodo de tiempo. La media de cada uno de los grupos estudiados es un estadístico muestral calculado a partir de los resultados del experimento.

Las poblaciones representadas por estas muestras ni siquiera existen realmente. Lo que sí existe es una población general de alumnos, por supuesto, pero no una población de alumnos a quienes se les hayan asignado las condiciones mencionadas (excepto en términos del experimento). Estamos interesados en analizar alumnos a los que en el futuro se les podrían dar tales instrucciones; se trata de una población desconocida. Generalmente, suponemos que estas poblaciones desconocidas están representadas por una curva normal, y lo hacemos simplemente porque la mayoría de las distribuciones en psicología lo están. Sin embargo, no tenemos ningún fundamento para realizar ninguna presunción sobre la media y la varianza de esas poblaciones: son parámetros poblacionales desconocidos. Cualquier conclusión que saquemos con respecto a los mismos debe basarse en la información proveniente de los estadísticos muestrales.

Finalmente, el tema que nos interesa es un tema relacionado con la probabilidad. El razonamiento es un poco complicado, por lo cual le dedicamos la mayor parte del capítulo 6. Sin embargo, para tener una noción previa del tema, analicemos la siguiente lógica: supongamos que las verdaderas medias de las dos poblaciones (parámetros poblacionales) fueran de hecho las mismas. Conforme a esta suposición, la *forma en que los alumnos estudien no* afecta el nivel de aprendizaje. No obstante, cuando realizamos el experimento, las calificaciones medias de los dos grupos en la prueba fueron diferentes. Entonces, dado nuestro supuesto de que no existe diferencia entre las poblaciones, ¿cuál es la probabilidad de que las medias de nuestras dos muestras pudieran ser tan diferentes como lo son en realidad? Si la probabilidad es baja, resulta poco verosímil que nuestro supuesto de que no existe diferencia entre las poblaciones sea correcto y, por lo tanto, lo rechazamos (el supuesto implica medias poblacionales iguales). Si rechazamos ese supuesto acerca de la inexistencia de diferencias entre las poblaciones, nos queda la conclusión de que existe diferencia entre las poblaciones. Es decir, este resultado sostiene la conclusión de que el modo de estudio de los estudiantes realmente afecta el nivel de aprendizaje.

La lógica que acabamos de describir puede parecer bastante intrincada, y de hecho lo es. Sin embargo, es justamente ese tipo de razonamiento sobre probabilidades, muestras y poblaciones el que fundamenta la mayor parte de la estadística inductiva en psicología. Es, en pocas palabras, la lógica de lo que se denomina "prueba de hipótesis", concepto que estudiaremos paso a paso en el capítulo 6. No necesitamos analizar ese tema ahora, sólo hemos introducido las ideas generales para dar una noción de la manera en que varios de los elementos tratados en este capítulo se combinan en los distintos tipos de problemas estadísticos que surgen de las investigaciones psicológicas reales.

CONTROVERSIAS Y LIMITACIONES

Aun siendo temas básicos, los tres conceptos presentados en este capítulo, la curva normal, la probabilidad y las muestras y poblaciones, son temas que generan bastante controversia. Analizaremos una importante controversia en relación con cada uno de ellos.

¿La curva normal, es realmente tan normal?

Hemos mencionado que las distribuciones reales con frecuencia se aproximan mucho al modelo de curva normal. Es muy importante saber hasta qué punto esto es verdad, y no sólo porque la presunción de modelo normal hace que las puntuaciones Z sean más útiles. Como veremos en capítulos posteriores, la mayoría de las técnicas estadísticas que utilizan los psicólogos suponen que sus muestras provienen de poblaciones distribuidas normalmente. El tema de en qué medida es razonable este supuesto ha sido una fuente de debate durante mucho tiempo. La postura predominante ha sido que, debido al modo en como se desarrollen las medidas psicológicas, la distribución con forma de campana "está prácticamente garantizada" (Walberg, Strykowski, Rovai, & Hung, 1984, p. 107). O, como lo expresaran Hopkins y Glass (1978), las mediciones en todas las disciplinas resultan ser tan buenas aproximaciones a ella que uno podría pensar "¡Dios ama la curva normal!"

Sin embargo, siempre ha existido una persistente postura crítica que plantea la pregunta de si la naturaleza en realidad se empaqueta tan prolijamente. Micceri (1989) presentó pruebas muy consistentes en el sentido de que muchas medidas comúnmente utilizadas en psicología no arrojan valores normalmente distribuidos "en la naturaleza". Su estudio incluía pruebas de nivel y capacitación (como el SAT y el GRE - *Graduate Record Examination*, Examen de inscripción de graduados), y pruebas de personalidad (como el MMPI - *Minnessota Multiphasic Personality Inventory*, Inventario de personalidad multifacética). Micceri obtuvo series de datos y analizó las distribuciones de los valores de 440 medidas psicológicas y educativas que habían sido observadas en muestras de gran tamaño. Todas sus series de datos correspondían a muestras de más de 190 individuos, y la mayoría correspondía a muestras de más de 1.000 (incluso un 14,3% correspondía a muestras de 5.000 a 10.293). Sin embargo, las muestras de gran tamaño no fueron muy útiles. Ninguna de las distribuciones investigadas pudo superar todas las pruebas de normalidad (Micceri buscaba fundamentalmente asimetrías, curtosis y "protuberancias"). Pocas medidas presentaban distribuciones que siquiera se acercaban razonablemente al modelo de la curva normal. Tampoco eran predecibles las variaciones: "Las distribuciones analizadas mostraron casi todos los casos concebibles de contaminación" (p. 162), aunque algunos eran más comunes en cierto tipo de pruebas. Micceri exhibe muchas razones obvias de esta anormalidad, tales como los efectos "piso" y "techo" (*véase* capítulo 2).

¿Qué importancia ha tenido el hecho de que las distribuciones de estas medidas fueran tan anormales? Según Micceri, simplemente se desconoce, y hasta que se sepa más sobre el tema, la opinión general entre los psicólogos continuará sosteniendo las técnicas estadísticas tradicionales, con la matemática implícita, que se basa en el supuesto de las distribuciones normales de población. ¿Cuál es la razón de esta indiferencia en vista de descubrimientos como los de Micceri? Sucede que en la mayoría de las condiciones en las que se las utiliza, las técnicas tradicionales parecen dar resultados razonablemente exactos, aun cuando no se cumpla el requerimiento formal de una distribución normal de población (p. ej. Sawilowsky & Blair, 1992). Este libro, en líneas generales, adopta la posición mayoritaria que favorece la utilización de técnicas tradicionales en todos los casos, excepto en los más extremos. Sin embargo, debemos tener en cuenta que existe una minoría resonante de psicólogos que están en desacuerdo con esto. En el capítulo 15 presentamos algunas de las técnicas estadísticas alternativas que esos psicólogos favorecen (técnicas que no están sustentadas por el supuesto de la distribución normal de las poblaciones).

Galton, uno de los pioneros más destacados en el campo de los métodos estadísticos (recordemos el cuadro 3-1), opinó sobre la curva normal: "No conozco casi nada tan apropiado para impresionar la imaginación [...] si los griegos hubieran sabido de ella la hubieran personificado y divinizado. Reina con serenidad y completa humildad en medio de la salvaje confusión" (1889, p. 66). Irónicamente, tal vez sea cierto que, al menos en psicología, realmente reina en un aislamiento puro y austero, sin imitaciones reales siquiera cercanas a lo perfecto.

¿Qué significa realmente la probabilidad?

Ya hemos presentado la mayor controversia con respecto a la teoría de la probabilidad, según se aplica a la estadística en psicología: el debate entre la interpretación como frecuencia relativa a largo plazo y la interpretación subjetiva según el grado de convencimiento. Sin embargo, en la mayoría de los casos, realmente no importa demasiado qué interpretación se utiliza, pues los cálculos estadísticos son los mismos. Pero entre la minoría de teóricos que favorecen la interpretación subjetiva, algunos sostienen una opinión bastante crítica de la rama principal del pensamiento estadístico. En particular, han defendido lo que ha devenido en llamarse "el método Bayesiano" (por ejemplo, *véase* Phillips, 1973). El método lleva el nombre de Thomas Bayes, un disidente clérigo inglés de principios del siglo XVIII, que desarrolló un teorema de la probabilidad adecuadamente denominado "Teorema de Bayes".

El teorema de Bayes puede ser probado matemáticamente, y no es controvertido. Sin embargo, sus aplicaciones en estadística son fuertemente discutidas. Los detalles del método exceden el alcance de un texto introductorio, pero sí podemos explicar claramente el principal tema en disputa: los bayesianos sostienen que la ciencia implica realizar investigaciones para adaptar nuestras creencias preexistentes a la luz de las pruebas recopiladas. Por lo tanto, las conclusiones derivadas de un experimento siempre se encuentran dentro del contexto de lo que creíamos sobre el mundo antes de realizar el experimento. La corriente principal, por el contrario, sostiene que es mejor no realizar ninguna presunción sobre creencias preexistentes. Deberíamos analizar las pruebas tal como son, juzgando si el experimento ha mostrado algún efecto confiable (o ningún tipo de efecto). Algunos estadísticos de la corriente principal reconocen que la descripción bayesiana de la ciencia puede ser más exacta. Sin embargo, no se sienten cómodos con la utilización de los métodos bayesianos en los cálculos estadísticos de las investigaciones prácticas, porque adoptarlos significaría que la conclusión obtenida a partir de cada estudio dependería demasiado de la creencia subjetiva del científico que está realizando el estudio. Y así, los mismos resultados experimentales podrían llevar a diferentes conclusiones si son analizados por diferentes científicos.

El método bayesiano representó un enérgico (aunque nunca mayoritario) movimiento en la estadística aplicada a la psicología durante las décadas de 1960 y 1970. Desde entonces se ha vuelto mucho menos prominente como movimiento, al menos bajo este estandarte. No obstante, muchas de las cuestiones que surgieron de esta disputa continúan siendo importantes bajo formas diferentes. (Games, 1988; Gigerenzer & Murray, 1987; Leventhal & Huyn, 1996; Prentice & Miller, 1992).

Muestra y población

La mayoría de los procedimientos estadísticos que aprenderemos en el resto de este libro se basan en el supuesto de que la muestra estudiada es una muestra aleatoria de la población. Como ya señalamos, sin embargo, esto rara vez sucede en la investigación psicológica. Lo más frecuente es que nuestras muestras incluyan a aquellos individuos que están disponibles para participar en un experimento, lo cual implica que la mayoría de los estudios se realicen con alumnos universitarios, voluntarios y animales de laboratorio que resulten convenientes y similares.

Algunos psicólogos se preocupan por este tema y han sugerido que los investigadores necesitan utilizar diferentes métodos estadísticos que realicen generalizaciones referidas sólo a los tipos de personas que en realidad están siendo utilizadas en el estudio. [3] Por ejemplo, estos psicólogos sostendrían que si nuestra muestra presenta una determinada distribución anormal, deberíamos suponer que se pueden generalizar los resultados sólo con respecto a una población con la misma distribución anormal. En el capítulo 15 seguiremos analizando estas sugerencias.

Los sociólogos, en comparación con los psicólogos, están mucho más preocupados por la representatividad del grupo que estudian. Es mucho más probable que se utilicen métodos formales de selección aleatoria y de grandes muestras en los estudios presentados en revistas especializadas en sociología (o en revistas científicas de psicología social orientadas a la sociología), o al menos que se trate el tema en sus publicaciones.

¿Por qué los psicólogos se sienten más cómodos utilizando muestras que no son claramente aleatorias? La razón más importante es que están interesados principalmente en las **relaciones** entre variables. Si en determinada población un aumento en *X* está relacionado con un aumento en *Y*, esa relación debería sostenerse probablemente en otras poblaciones, y debería hacerlo incluso si los niveles reales de *X* e *Y* son diferentes entre las poblaciones. Supongamos que un investigador realiza el experimento que utilizamos como ejemplo en el capítulo 3 y 4, probando la relación entre la cantidad de veces que se expone una lista de palabras con la cantidad de palabras recordadas. Supongamos, además, que el estudio se realiza con alumnos universitarios, y que el resultado es que, a mayor cantidad de exposiciones, mayor cantidad de palabras recordadas. La cantidad real de palabras recordadas de la lista bien podría ser diferente, en el caso de personas pertenecientes a grupos sociales distintos, al de los alumnos universitarios. Por ejemplo, es probable que expertos en ajedrez (quienes probablemente tengan la memoria altamente desarrollada) puedan recordar más palabras; personas que acaban de sufrir algún trastorno probablemente recuerden menos palabras. Sin embargo, incluso en esos grupos, esperaríamos que, a mayor cantidad de exposiciones de la lista, más palabras fueran recordadas. Por lo tanto, es probable que la **relación** entre cantidad de exposiciones y cantidad de palabras recordadas sea aproximadamente la misma en cada población.

En sociología, la representatividad de las muestras es mucho más importante debido a que los sociólogos están más preocupados por la media y la varianza real de una variable en determinada sociedad. Así, un sociólogo podría estar interesado en la actitud promedio hacia las personas mayores en la población de un determinado país. En ese caso, es extremadamente importante la manera en que se realice el muestreo.

[3] Frick (en prensa) sostiene que en la mayoría de los casos los investigadores psicológicos no deberían pensar siquiera en función de muestras y poblaciones, sino que más bien deberían considerarse investigadores estudiando procesos. Un experimento analiza algún proceso en un grupo de individuos. Luego, el investigador evalúa la probabilidad de que el patrón de resultados pudiera haber sido causado por factores casuales. Por ejemplo, el investigador analiza si una diferencia de medias entre un grupo experimental y uno de control podría haber sido causada por otros factores además de la manipulación experimental. Frick sostiene que este modo de pensar es mucho más parecido a la forma real en que los investigadores trabajan, y afirma que presenta varias ventajas en cuanto a la sutil lógica de los procedimientos de estadística inductiva. Será interesante ver la reacción a la propuesta de Frick. En todo caso, seguir el método más estándar (tal como se enseña en este libro) arroja exactamente los mismos resultados, lo cual es coherente con la manera en la que la mayoría de los psicólogos comprenden el razonamiento estadístico.

CURVAS NORMALES, PROBABILIDADES, MUESTRAS Y POBLACIONES SEGÚN SE DESCRIBEN EN PUBLICACIONES CIENTÍFICAS

Los temas tratados en este capítulo se utilizan especialmente como base para comprender el material expuesto en los capítulos siguientes, y rara vez se nombran explícitamente en publicaciones científicas (excepto en artículos sobre métodos o cálculos estadísticos). Ocasionalmente, podremos ver que se menciona la curva normal en el contexto de la descripción de valores de una determinada variable. (En el capítulo 15 proporcionamos más información acerca de este tema, e incluso algunos ejemplos tomados de publicaciones reales. En ese mismo capítulo también analizamos circunstancias en las que los valores no siguen la distribución normal).

Tampoco es común que se mencione la probabilidad de manera directa, excepto en el contexto de la significación estadística, tema que mencionamos brevemente en el capítulo 3. En casi cualquier publicación que tengamos la oportunidad de leer, la sección "Resultados" estará llena de descripciones de distintos métodos relacionados con la significación estadística, seguidas de expresiones tales como "$p<0{,}05$" ó "$p<0{,}01$". La p se refiere a probabilidad, pero, ¿probabilidad de qué? Ese es el tema principal de nuestra exposición sobre significación estadística en el capítulo 6.

Finalmente, sólo en algunas ocasiones encontraremos una breve mención del método utilizado para seleccionar la muestra de la población. Por ejemplo, Altman, Levine, Howard y Hamilton (1997) realizaron una encuesta telefónica sobre las actitudes del público adulto norteamericano hacia los agricultores de tabaco. En la sección del artículo dedicada al método, explican que los que respondieron fueron "seleccionados en forma aleatoria de una lista nacional de números telefónicos" (p. 117). Así, Altman et. al. especificaban tanto la lista que utilizaron para elegir la población (el directorio nacional de números telefónicos) como el método utilizado (selección aleatoria) para obtener la muestra. Cabe destacar, sin embargo, que en tales encuestas el porcentaje de respuesta de aquellos a quienes se llama por teléfono generalmente es muy lejano al 100%. En el ejemplo que estamos analizando, obtuvieron entrevistas con el 47% de las personas a las que llamaron. Por lo tanto, aunque utilizaron el método de selección aleatoria para contactar a miembros potenciales de su muestra, la muestra propiamente dicha no fue aleatoria. La muestra representa excesivamente cualesquiera sean las características que hacen que una persona esté disponible y dispuesta a responder una encuesta telefónica.

RESUMEN

En muchas de las variables analizadas en la investigación psicológica, la distribución de los valores presenta aproximadamente una forma de campana, simétrica y unimodal, a la que llamamos curva normal. Dado que la forma de esta curva responde a una fórmula matemática exacta, existe un porcentaje específico de valores entre cualesquiera dos puntos de ella.

Las cifras importantes que conviene recordar con respecto a una curva normal son: un 34% de los valores se encuentran entre la media y 1 desvío estándar por encima de la media, y un 14% entre 1 y 2 desvíos estándar por encima de ella.

Una tabla de áreas de la curva normal indica el porcentaje de valores entre la media y cualquier puntuación Z positiva en particular. Utilizando esa tabla, y sabiendo que la curva es simétrica y que el 50% de los valores se encuentran por encima de la media, podemos determinar el porcentaje de valores por encima o por debajo de cualquier puntuación Z en particular. También podemos utilizar la tabla para determinar la puntuación Z correspondiente al punto en el que comienza un determinado porcentaje de valores.

La mayoría de los investigadores psicológicos considera que la probabilidad de un hecho es su frecuencia relativa esperada. Sin embargo, algunos consideran a la probabilidad como el grado subjetivo de convencimiento de que el hecho sucederá. La probabilidad generalmente se calcula como la razón entre la cantidad de resultados favorables y la cantidad total de resultados posibles. Se simboliza con una p y tiene un rango de 0 (hecho imposible) a 1 (hecho cierto). El área bajo la curva normal indica la probabilidad de que los valores se ubiquen dentro de determinado intervalo de valores.

Una muestra es un individuo o grupo analizado, por lo general en representación de un grupo mayor o población que no puede ser analizado en su totalidad. Lo ideal es que la muestra sea seleccionada de la población utilizando un procedimiento estrictamente aleatorio. La media, la varianza y demás cálculos de una muestra se denominan estadísticos muestrales. Cuando se refieren a una población, se denominan parámetros poblacionales y se simbolizan con letras griegas(μ, para la media, σ^2 para la varianza y σ para el desvío estándar).

La mayoría de las técnicas que aprenderemos en el resto del libro utilizan inferencias probabilísticas para sacar conclusiones acerca de poblaciones, sobre la base de información obtenida a partir de muestras. En este proceso, generalmente se presume que las poblaciones están normalmente distribuidas.

Existen controversias con respecto a cada uno de los temas principales. Una de las cuestiones se refiere a si las distribuciones normales son realmente típicas de las poblaciones de valores correspondientes a las variables que estudiamos en psicología. Otro debate, planteado por defensores del enfoque "bayesiano" de la estadística, es si deberíamos construir explícitamente los procedimientos estadísticos de forma tal de tener en cuenta las expectativas subjetivas iniciales del investigador. Finalmente, se ha discutido la representatividad de las muestras utilizadas por los psicólogos, que en líneas generales no se obtienen a través de una selección estrictamente aleatoria, aunque existen también motivos para pensar que con respecto a los temas que estudian la mayoría de los psicólogos, este punto no tiene gran relevancia.

Las publicaciones científicas rara vez exponen las curvas normales (excepto brevemente cuando la distribución que se está analizando parece no ser normal) o la probabilidad (excepto en el contexto de las pruebas de significación, descriptas al comienzo del capítulo 6). Sin embargo, en líneas generales sí se describen los procedimientos de muestreo, especialmente cuando el estudio es un sondeo de datos; y se puede discutir la representatividad de una muestra cuando no hubiera sido posible realizar un muestreo al azar.

Términos clave

- Frecuencia relativa esperada.
- Selección casual.
- Interpretación de la probabilidad como la frecuencia relativa a largo plazo.
- Curva normal.
- Tabla de áreas de la curva normal.
- Distribución normal.
- Resultado.
- Población.
- Parámetros poblacionales.
- Probabilidad (p).
- Selección aleatoria.
- Muestra.
- Estadísticos muestrales.
- Interpretación subjetiva de probabilidad.
- μ.
- σ.
- σ^2

Ejercicios

Los ejercicios implican la realización de cálculos (con la ayuda de una calculadora). La mayoría de los problemas estadísticos reales se resuelven por computadora, pero aunque exista la posibilidad de utilizar una computadora, es conveniente realizar estos ejercicios manualmente para incorporar el método de trabajo.

Para adquirir práctica en la utilización de una computadora, para resolver problemas estadísticos, se puede utilizar la sección de computación de cada capítulo, publicada en la *Guía de estudio y libro de tareas de computación para el alumno [Student's Study Guide and Computer Workbook]* que acompaña este libro.

Todos los datos de esta sección son ficticios (a menos que se especifique lo contrario).

Las respuestas a los ejercicios de la serie I se encuentran al final del libro.

SERIE I

1. Supongamos que las personas que viven en determinada ciudad tuvieran una media de 40 y un desvío estándar de 5 con respecto a la preocupación sobre el medio ambiente. Suponiendo que estos valores referidos a la preocupación están normalmente distribuidos: ¿Qué porcentaje aproximado de personas presenta un registro a) mayor a 40, b) mayor a 45, c) mayor a 30, d) mayor a 35, e) menor a 40, f) menor a 45, g) menor a 30 y h) menor a 35? ¿Cuál es el valor mínimo que una persona debe tener para estar en el i) 2%, j) 16%, k) 50%, l) 84%, y m) 98% superior? (Utilice los números 50%–34%–14% para resolver este problema).

2. Una psicóloga estudió la fatiga ocular utilizando una medida particular que aplica a los alumnos después de 1 hora de trabajo escribiendo en una computadora. Con esta medida, la psicóloga ha descubierto que la distribución presenta una curva normal. ¿Qué porcentaje de alumnos presenta una puntuación Z a) menor a 1,5, b) mayor a 1,5, c) menor a –1,5, d) mayor a –1,5, e) mayor a 2,10, f) menor a 2,10, g) mayor a 0,45, h) menor a –1,78 y i) mayor a 1,68?

3. Suponiendo que se trata de una distribución normal, a) si una persona se encuentra entre el 10% superior de su país en cuanto a capacidad matemática, ¿cuál es la puntuación Z de esa persona? b) Si la persona se encuentra dentro del 1%, ¿cuál sería la puntuación Z?

4. Analicemos una prueba de coordinación con distribución normal, una media de 50 y un desvío estándar de 10. ¿Qué valor necesitaría una persona para estar entre el 5% superior? Explique su respuesta a alguien que nunca ha tomado un curso de estadística.

5. Las siguientes cantidades de individuos de una empresa recibieron atención especial de la gerencia de personal el año pasado:

Drogas/alcohol	10
Asesoramiento para crisis familiar	20
Varios	20
Total	50

Si de los expedientes del año pasado tuviera que seleccionar a alguien al azar, ¿cuál es la probabilidad de que la persona sea del grupo a) drogas / alcohol, b) familia, c) drogas / alcohol o familia, d) cualquier categoría excepto "Varios", e) cualquiera de las tres categorías?

6. Una publicación científica trata el tema del nivel de autoestima de los alumnos secundarios australianos. En la sección en la que se describen los métodos aplicados se remarca que se estudió una "muestra aleatoria" de alumnos secundarios. Explique a una persona que nunca ha realizado un curso de estadística o métodos de investigación qué significa esto y por qué es importante.

SERIE II

1. Se descubre que la cantidad de tiempo que toma recuperarse fisiológicamente de determinado estímulo está distribuida normalmente con una media de 80 segundos y un desvío estándar de 10 segundos. ¿Aproximadamente qué porcentaje de registros (en tiempo de recu-

peración) estará a) por encima de 100, b) por debajo de 100, c) por encima de 90, d) por debajo de 90, e) por encima de 80, f) por debajo de 80, g) por encima de 70, h) por debajo de 70, i) por encima de 60 y j) por debajo de 60? ¿Cuál es el periodo de tiempo más largo que puede tardar una persona en recuperarse y aun así pertenecer al k) 2%, l) 16%, m) 50%, n) 84% y o) 98% inferior? (Utilice los números 50%–34%–14% para resolver este problema).

2. Supongamos que las puntuaciones de arquitectos en determinada prueba de creatividad están distribuidas normalmente. ¿Qué porcentaje de arquitectos tiene puntuaciones Z a) mayores a 0,10, b) menores a 0,10, c) mayores a 0,20, d) menores a 0,20, e) mayores a 1,10, f) menores a 1,10, g) mayores a –0,10, y h) menores a –0,10?

3. En el caso del problema 2, ¿cuál es la puntuación Z mínima que puede tener un arquitecto en la prueba de creatividad para estar dentro del a) 50% superior, b) 40% superior, c) 60% superior, d) 30% superior y e) 20% superior?

4. Supongamos que está diseñando un panel de instrumentos para una gran máquina industrial que requiere un alcance de 2 pies desde determinada posición. Se sabe que el alcance desde esa posición para mujeres adultas presenta una media de 2,8 pies, con un desvío estándar de 0,5. El alcance para hombres adultos presenta una media de 3,1 pies con un desvío estándar de 0,6. Tanto el alcance de las mujeres como de los hombres desde esa posición está normalmente distribuido. Si se implementa este diseño, ¿qué porcentaje de mujeres no podrán trabajar con ese panel de instrumentos? ¿Qué porcentaje de hombres no podrá trabajar con ese panel de instrumentos? Explique sus respuestas a una persona que nunca ha tomado un curso de estadística.

5. Se realiza una encuesta en una facultad con 800 alumnos, 50 miembros del claustro docente y 150 empleados administrativos. Cada uno de estos 1.000 individuos aparece una sola vez en el directorio telefónico del campo universitario. Supongamos que tuviera que abrir el directorio y sacar un número al azar para contactar. ¿Cuál es la probabilidad de que sea a) un alumno, b) un miembro del claustro docente, c) un miembro del personal administrativo, d) un miembro del claustro docente o un empleado administrativo y e) cualquiera, excepto un miembro del claustro docente o personal administrativo?

6. Supongamos que usted fuera a realizar una encuesta a visitantes de su campo universitario y quiere que la encuesta resulte lo más representativa posible. ¿Cómo seleccionaría las personas a encuestar? ¿Por qué considera que ese sería el mejor método?

APÉNDICE DEL CAPÍTULO: REGLAS DE LA PROBABILIDAD Y PROBABILIDADES CONDICIONALES

Las reglas de la probabilidad son procedimientos para calcular probabilidades que involucran múltiples experimentos o resultados. Las dos reglas más ampliamente utilizadas son la **regla de la adición** (también llamada la **regla o**) y la **regla de la multiplicación** (también llamada la **regla** *y*). En este apéndice analizamos esas dos reglas y explicamos también el concepto de probabilidades condicionales.

Regla de la *adición*

La regla de la adición se aplica a situaciones que involucran la probabilidad de obtener cualquiera de dos o más **resultados mutuamente excluyentes**. Resultados mutuamente excluyentes son aquellos en los que la ocurrencia de un resultado hace que el otro resultado no suceda, como pueden ser los resultados cara o ceca en un sólo tiro de monedas o los resultados de uno o seis en un sólo tiro de un dado. En el caso de los resultados mutuamente excluyentes, la probabilidad de obtener algunos de ellos es la suma de las probabilidades individuales. Por lo tanto, en un sólo tiro de moneda, las posibilidades de obtener cara (que es de 0,5) o ceca (también de 0,5) es de 1,0 (0,5 más 0,5). En un sólo tiro de un dado, las posibilidades de obtener un 3 (1/6) ó un 5 (1/6) son de *1/3 (1/6 +1/6). Si* usted elige a un alumno de su universidad al azar, y en su universidad un 30% de los alumnos son avanzados y un 25% son principiantes, la posibilidad de elegir a alguien que sea avanzado o principiante es del 55%.

La regla formalmente se expresa:

$$p(\text{A ó B}) = p(\text{A}) + p(\text{B}) \tag{5-1}$$

Donde p(A ó B) es la probabilidad de obtener el resultado A o el resultado B; p(A) es la probabilidad de obtener el resultado A, y p(B) es la probabilidad de obtener el resultado B.

La regla de adición se aplica cualquiera sea la cantidad de resultados mutuamente excluyentes. Por ejemplo, $p(\text{A, B, ó C}) = p(\text{A}) + p(\text{B}) + p(\text{C})$.

Regla de la multiplicación

La regla de la multiplicación se aplica a situaciones que involucran más de un experimento. Permite calcular la probabilidad de obtener **ambos** de dos (o más) **resultados independientes**. Los resultados independientes son tales que el acontecimiento de uno no da al otro mayor ni menor probabilidad de suceder. Obtener cara o ceca en un tiro de moneda es un resultado independiente de obtener cara o ceca en un segundo tiro de moneda. La probabilidad de obtener ambos de los dos resultados independientes es el producto de (el resultado de multiplicar) las probabilidades individuales. Por ejemplo, en un sólo tiro de moneda, la posibilidad de obtener cara es de 0,5. En un segundo tiro de moneda, la probabilidad de obtener cara (sin importar lo que se obtuvo en el primer tiro) es también de 0,5. Por lo tanto, la probabilidad de obtener caras en **ambos** tiros de moneda es de 0,25 (0,5 por 0,5). En dos tiros de un dado, la probabilidad de obtener un 5 en ambos tiros es igual a 1/36, es decir, la probabilidad de obtener un 5 en el primer tiro (1/6), multiplicada por la probabilidad de obtener un 5 en el segundo tiro (1/6). De modo similar, en una prueba de **selección múltiple** con cuatro opciones para cada ítem, la probabilidad de adivinar dos respuestas correctas es de 1/16, es decir, la probabilidad de adivinar una respuesta correcta (1/4) multiplicada por la posibilidad de adivinar la otra respuesta correcta (1/4).

Expresado por una fórmula:

$$p(\text{A y B}) = p(\text{A}) \times p\,(\text{B}) \tag{5-2}$$

Donde p(A y B) es la probabilidad de obtener el resultado A y el resultado B (suponiendo que son resultados independientes).

La regla de la multiplicación se aplica cualquiera sea la cantidad de resultados independientes. Por ejemplo, $p(\text{A, B y C}) = p(\text{A}) \times p(\text{B}) \times p(\text{C})$.

Probabilidades condicionales

Existen otras reglas de la probabilidad (algunas de las cuales son combinaciones de las mencionadas anteriormente). La mayoría involucra lo que se denomina **probabilidades condicionales.**

Una probabilidad condicional es la probabilidad de un resultado suponiendo que otro resultado ha ocurrido. Es decir, la probabilidad de un resultado está condicionada por la ocurriencia del otro resultado. Por lo tanto, supongamos que la facultad A tiene un 50% de mujeres y la facultad B tiene un 60% de mujeres. Si seleccionamos una persona al azar, ¿cuál es la posibilidad de que resulte seleccionada una mujer? Si sabemos que la persona es de la facultad A, la probabilidad es del 50%. Es decir, la probabilidad de que resulte seleccionada una mujer, a condición de que provenga de la facultad A, es del 50%. Diríamos que, *p*(mujer / facultad A) =0,5. De modo similar, *p*(mujer / facultad B) = 60%.

6 Introducción a la prueba de hipótesis

Descripción del capítulo

- Un ejemplo de prueba de hipótesis.
- Lógica central de la prueba de hipótesis.
- El proceso de la prueba de hipótesis.
- Pruebas de hipótesis de una y dos colas.
- Controversias y limitaciones.
- La prueba de hipótesis según se describe en las publicaciones científicas.
- Resumen.
- Términos clave.
- Ejercicios.

En el capítulo 5 aprendimos los conceptos de curva normal, probabilidad y la diferencia entre una muestra y una población. En este capítulo, presentamos el tema crucial de la prueba de hipótesis. La prueba de hipótesis es un procedimiento sistemático para determinar si los resultados de un experimento a través del cual se analiza una muestra, sustentan una teoría o innovación práctica determinada que se aplica a una población. La **prueba de hipótesis** es el tema central de todos los capítulos restantes de este libro, como lo es también en la mayoría de las investigaciones científicas. Casi todos las publicaciones de investigación psicológica utilizan la prueba de hipótesis.

Es nuestro deber advertir que, para la mayoría de los alumnos, la parte más difícil del curso es el manejo de la lógica básica de este capítulo y de los dos siguientes. Este capítulo, en particular, requiere cierta gimnasia mental. Aun cuando se comprendan todos los razonamientos la primera vez, es recomendable realizar una revisión completa. La prueba de hipótesis involucra un grupo de ideas que, contempladas separadamente, no tienen mucho sentido. Por lo tanto, en este capítulo aprenderemos una cantidad comparativamente grande de ideas al mismo tiempo. Mirando el lado positivo, una vez que hayamos incorporado bien los temas de este capítulo y de los dos siguientes, estaremos acostumbrados a este tipo de material, y el resto del curso resultará sencillo.

Al mismo tiempo, hemos desarrollado esta introducción a la prueba de hipótesis de la manera más sencilla posible, y dejamos para los capítulos posteriores todo aquello que podía postergarse. Por ejemplo, las investigaciones psicológicas reales casi siempre involucran muestras compuestas por muchos –a veces muchísimos– individuos. Sin embargo, para simplificar las cosas, todos los ejemplos de este capítulo se refieren a estudios en los que la muestra está formada por un sólo individuo. Para lograrlo, hemos tenido que crear algunos ejemplos bastante extraños, por eso es conveniente que el alumno recuerde simplemente que estamos construyendo los cimientos que, en el capítulo 9, lo prepararán para comprender la prueba de hipótesis tal como se realiza en la realidad.

UN EJEMPLO DE PRUEBA DE HIPÓTESIS

Este es el primer ejemplo ficticio y, necesariamente, extraño. Durante varios años se ha desarrollado un gran proyecto de investigación. En el contexto del proyecto, se ha administrado a bebés recién nacidos una vitamina especial, y luego se ha controlado su desarrollo durante los primeros dos años de vida. Hasta ahora, la vitamina no ha acelerado el desarrollo de los bebés. La distribución de la edad en la que éstos y todos los bebés comienzan a caminar está representada por la figura 6-1. En ella observamos que la media es 14 meses, el desvío estándar es de 3 meses, y las edades siguen una curva normal. Mirando la curva podemos observar que menos del 2% de los bebés comienzan a caminar antes de los 8 meses de edad (estos bebés se encuentran 2 desvíos estándar por debajo de la media de edad para comenzar a caminar). (La distribución que analizamos, si bien es ficticia, en realidad es bastante similar a la distribución que los psicólogos han probado en el caso de bebés europeos, aunque esa distribución real es levemente asimétrica hacia la derecha; Hindley, Filliozat, Klackenberg, Nicolet-Meister, & Sand, 1966).

Uno de los investigadores del proyecto ha tenido una idea. Sobre la base de algunas nuevas teorías, razona que si la vitamina que toman los bebés estuviera más refinada, su efecto podría ser notablemente mayor, y que los bebés que tomaran la versión con alto grado de refinamiento deberían comenzar a caminar mucho antes que los otros bebés. (Supondremos que el proceso de purificación no podía de ningún modo hacer que la vitamina fuera dañina para los bebés). Sin embargo, refinar la vitamina de este modo eleva en gran medida el costo de cada dosis; por lo tanto, el equipo de investigación decide probar el procedimiento con dosis suficientes para un sólo bebé. Entonces, se selecciona al azar un bebé del proyecto para suministrarle la versión altamente refinada de la vitamina, y se realiza un seguimiento de su progreso junto con el de todos los otros bebés del mismo proyecto. ¿Qué tipo de resultado llevaría a los investigadores a sacar la conclusión de que la vitamina altamente purificada hace que los bebés caminen a más temprana edad?

Lo que acabamos de describir es el ejemplo de un problema que se resuelve a través de la prueba de hipótesis. Los investigadores pretenden sacar una conclusión acerca de si la vitamina purificada hace que los bebés en general caminen antes de lo esperado. La conclusión referida a los bebés en general, sin embargo, se basará en los resultados obtenidos, estudiando sólo una muestra. (En este extraño ejemplo, la muestra es un sólo bebé).

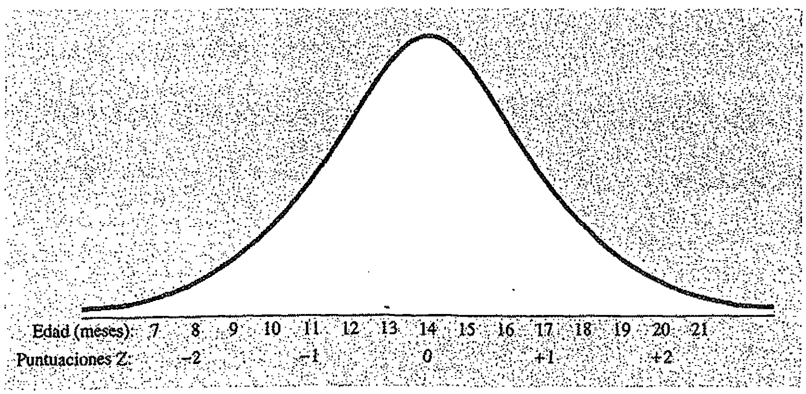


Figura 6-1. Distribución de edades en que los bebés comienzan a caminar (datos ficticios).

LÓGICA CENTRAL DE LA PRUEBA DE HIPÓTESIS

Existe un método estándar para encarar un problema de prueba de hipótesis. El investigador utilizará el siguiente razonamiento: comúnmente, las chances de que un bebé comience a caminar a los 8 meses de edad o antes serían menores al 2%. Por lo tanto, caminar a los 8 meses es altamente inverosímil. ¿Pero qué sucede si el bebé que estamos estudiando comienza a caminar a los 8 meses? Si esto sucede, podremos **rechazar** la idea de que la vitamina especialmente purificada **no** produce ningún efecto. Si rechazamos la idea de que la vitamina especialmente purificada no produce ningún efecto, debemos **aceptar** la idea de que sí produce un efecto. (La lógica de este ejemplo es crucial para todos los siguientes temas del libro. Tal vez sea conveniente volver a leer este párrafo).

En primer lugar, los investigadores han comprendido qué tendría que suceder para poder sacar la conclusión de que el procedimiento de purificación especial marca una diferencia. Habiendo comprendido esto previamente, los investigadores pueden entonces continuar con la realización de su estudio. En este caso, realizar el estudio significa suministrar la vitamina especialmente purificada a un determinado bebé y observar a qué edad ese bebé comienza a caminar. Si el resultado del estudio muestra que el bebé comienza a caminar antes de los 8 meses, entonces concluirán que es inverosímil que la vitamina especialmente purificada no provoque una diferencia. Si es inverosímil que la vitamina especialmente purificada no provoque una diferencia, entonces la conclusión es que probablemente sí la provoque.

Este tipo de razonamiento al revés, contrario a lo que uno predice, es el corazón de la estadística inferencial en psicología. Es algo así como una doble negación. Uno de los fundamentos de este método es que podemos determinar directamente la probabilidad de obtener un resultado experimental determinado si la situación de que no se produzca diferencia es verdadera. En el ejemplo de la vitamina purificada, los investigadores saben cuáles son las probabilidades de que los bebés caminen a diferentes edades si la vitamina especialmente purificada no produce ningún efecto. Es la probabilidad de que un bebé camine a distintas edades lo que ya conocemos por analizar bebés en general, es decir, bebés que no han recibido vitamina especialmente purificada. (Supongamos que la vitamina especialmente purificada no produce ningún efecto. En ese caso, la edad en la que los bebés comienzan a caminar es la misma, reciban o no la vitamina especialmente purificada. Por lo tanto, la distribución es la que aparece en la figura 6-1, basada en las edades en las que los bebés en general comienzan a caminar).

Sin esta reconocidamente tortuosa manera de enfocar el problema, en la mayoría de los casos no habría modo de probar una hipótesis. En casi todas las investigaciones psicológicas, ya sea con experimentos, encuestas u otro método, sacamos conclusiones evaluando la probabilidad de obtener nuestros resultados de investigación si fuera verdad lo contrario a lo que estamos prediciendo. Es decir, generalmente predecimos algún tipo de efecto pero evaluamos si existe tal efecto observando si es inverosímil la hipótesis de que ese efecto no exista.

EL PROCESO DE LA PRUEBA DE HIPÓTESIS

Volveremos a analizar la solución del problema de prueba de hipótesis que utilizamos como ejemplo estudiando cada paso con mayor detalle, así como también algunos de los términos especiales que se han utilizado. Al hacerlo, presentaremos un procedimiento de cinco pasos que se utilizará en el resto del libro.

1° paso: replantear el problema sobre las poblaciones en función de hipótesis de investigación e hipótesis nula

Primero, tengamos en cuenta que los investigadores están interesados en los efectos provocados en los bebés en general (no sólo en el bebé en particular que ellos estudian). Por lo tanto, será útil volver a plantear el problema en función de poblaciones. Con el propósito de analizar esta situación, podemos decir que los bebés se dividen en dos grupos:

Población 1: bebés que toman la vitamina especialmente purificada.
Población 2: bebés que no toman la vitamina especialmente purificada.

La población 1 se refiere a aquellos que reciben el tratamiento experimental. En nuestro ejemplo, existe sólo un caso real de población 1. Sin embargo, ese único bebé representa un futuro grupo de muchos bebés que aún no han nacido, y a quienes los investigadores pretenden aplicar sus resultados. La población 2 representa una especie de línea de base de lo conocido.

La predicción del investigador está basada en una teoría acerca de cómo funcionan las vitaminas de este tipo. La predicción es que los bebés de la población 1 (aquellos que toman la vitamina especialmente purificada) en general caminarán antes que los de la población 2 (aquellos que no toman la vitamina especialmente purificada). Una afirmación de este tipo, acerca de la diferencia entre poblaciones predichas sobre la base de una teoría (o basada en la experiencia práctica), se denomina **hipótesis de investigación**. Para decirlo de modo más concreto, decimos que la predicción establece que la media de la población 1 es menor (los bebés que reciben la vitamina especial caminan antes) que la media de la población 2. En símbolos, la hipótesis de investigación es $\mu 1 < \mu 2$.

¿Qué sucede si la predicción es incorrecta? En ese caso, se mantiene la situación contraria: los bebés de la población 1 (aquellos que toman la vitamina especialmente purificada) en general no caminarán antes que los bebés de la población 2 (aquellos que no toman la vitamina especialmente purificada). Esta predicción opuesta implica que no existe diferencia en cuanto al momento en que los bebés de la población 1 y la población 2 comienzan a caminar, es decir, comienzan al mismo tiempo. Una afirmación de este tipo, acerca de la ausencia de diferencia entre poblaciones, es el punto crítico **opuesto** a la hipótesis de investigación. Se denomina **hipótesis nula** porque se utiliza generalmente para indicar una situación en la que no existe diferencia entre dos poblaciones (la diferencia es nula). En símbolos, la hipótesis nula es $\mu 1 = \mu 2$.[1]

La hipótesis de investigación y la hipótesis nula son completamente opuestas. Si una es verdadera, la otra no puede serlo. Esta oposición, y la concentración directa en la hipótesis nula, es un punto central de la lógica de la prueba de hipótesis. Por ese motivo, la hipótesis de investigación, que finalmente es lo que realmente nos interesa, con frecuencia se denomina "hipótesis alternativa". En realidad, la situación es un poco irónica. Desde el punto de vista de nuestro interés en el asunto, lo que más nos importa es la hipótesis de investigación. Sin embargo, desde el punto de vista de la prueba de hipótesis, el papel principal de la hipótesis de investigación es su condición de alternativa de la hipótesis nula.

[1] En este caso hemos simplificado el tema. La hipótesis de investigación implica que una población caminará antes que la otra, $\mu_1 < \mu_2$. Por lo tanto, lo contrario implica que el otro grupo caminará o bien al mismo tiempo o después. Así, lo contrario a la hipótesis de investigación, en este caso incluye tanto la falta de diferencia como una diferencia en dirección contraria a la predicha. En términos de símbolos, si nuestra hipótesis de investigación es $\mu_1 < \mu_2$, entonces su opuesto es $\mu_1 \geq \mu_2$ (el símbolo $\geq$ significa "mayor o igual a"). Presentamos este tema con mayor detalle más adelante en este capítulo. Por ahora, para simplificar el aprendizaje, algunas veces consideraremos que la hipótesis nula implica que las dos poblaciones son esencialmente iguales, y otras veces consideraremos que implica que una población es igual u opuesta a la hipótesis de investigación.

2° paso: determinar las características de la distribución comparativa

Una vez que hemos planteado la situación en términos de elección entre una hipótesis de investigación y una hipótesis nula, el siguiente paso es analizar cómo podríamos utilizar la información que obtenemos sobre una muestra para realizar esta elección. La pregunta que planteamos es la siguiente: dado un determinado resultado muestral (en este caso, una observación), ¿qué probabilidad teníamos de obtener ese resultado si la hipótesis nula fuera verdadera?

Para responder esta pregunta, debemos saber cómo sería la situación si la hipótesis nula fuera verdadera. Es decir, necesitamos conocer los detalles de la distribución de la población de la cual proviene la muestra si la hipótesis nula fuera verdadera. Si conocemos la distribución de la población de la que proviene nuestra muestra, y sabemos que se trata de una distribución normal, nos encontramos en una buena posición: podemos determinar directamente la probabilidad de obtener cualquier valor determinado de esa distribución utilizando una tabla de áreas bajo la curva normal.

¿Cómo podemos conocer los detalles de la población de la cual proviene nuestra muestra si la hipótesis nula es verdadera? Esto es posible porque, si la hipótesis nula es verdadera, ambas poblaciones son iguales. Generalmente conocemos una de las poblaciones (población 2); por lo tanto, si la hipótesis nula es verdadera y las dos poblaciones son iguales, también conocemos la otra población (población 1). En nuestro ejemplo, si la hipótesis nula es verdadera, ambas poblaciones siguen la curva normal, y presentan una media de 14 meses y un desvío estándar de 3 meses (*véase* figura 6-1).

En este libro llamaremos a la distribución correspondiente a la situación en la que la hipótesis nula es verdadera, es decir, la distribución con la que comparamos la muestra, **distribución comparativa**. (La distribución comparativa a veces es denominada "modelo estadístico", y en la mayoría de los casos también coincide con lo que se denomina una "distribución muestral", una idea que expondremos en el capítulo 7). Es decir, en el proceso de la prueba de hipótesis, comparamos los valores observados en la muestra con esta distribución. Realizamos la comparación calculando la probabilidad de obtener un valor tan extremo como el de nuestra muestra en esa distribución comparativa. En el ejemplo que estamos tratando, la distribución comparativa es igual a la distribución de valores de la población 2, la población a la que no se le ha aplicado el procedimiento experimental.

3° paso: determinar el valor muestral de corte en la distribución comparativa, en el que debería rechazarse la hipótesis nula

Lo ideal sería que antes de realizar un estudio, los investigadores establezcan un objetivo con el cual comparar su resultado, es decir, qué valor extremo necesitaría tener la muestra para poder sacar una conclusión confiable. Específicamente, determinar el valor que necesitaría arrojar la muestra para decidir rechazar la hipótesis nula, cuán extremo debería ser ese valor para que resulte demasiado improbable que pudiera obtenerse tal valor extremo si la hipótesis nula fuera verdadera. A esto se lo denomina **punto muestral de corte** (también se conoce con el nombre de "punto crítico").

Analicemos nuestro ejemplo de la vitamina purificada, en el que la hipótesis nula implica que no importa si un bebé recibe la vitamina especialmente purificada o no. Los investigadores podrían decidir que si la hipótesis nula fuera verdadera, sería muy improbable que un bebé camine a los 8 meses o antes. Tratándose de dos desvíos estándar por debajo de la media (caminar a los 8 meses), sólo podría ocurrir menos de un 2% de las veces. Por lo tanto, basándose en la distribución comparativa, los investigadores establecen su punto muestral de corte incluso antes de realizar el estudio. Lo que están haciendo es decidir por adelantado que, si el resultado de su estudio es un bebé que camina antes de los 8 meses, rechazarán la hipótesis nula.

Si el bebé comienza a caminar antes de los 8 meses, los investigadores podrán rechazar la hipótesis nula, y si lo hacen, les quedará la hipótesis de investigación. Entonces podrán decir que "se sostiene la hipótesis de investigación".

Por otro lado, si el bebé no comienza a caminar sino hasta después de los 8 meses, no podrán rechazar la hipótesis nula. Cabe notar, sin embargo, que en este caso no podrán decir "se sostiene la hipótesis nula", ya que no rechazar la hipótesis nula crea una situación ambigua. No pueden sacarse conclusiones, excepto, concluir que se necesita realizar una mayor investigación. Más adelante continuaremos tratando este tema.

Al establecer por adelantado cuán extremo deberá ser un valor para rechazar la hipótesis nula, los investigadores no utilizan por lo general un número real de unidades de la escala directa de medición (en este caso, meses). En cambio, establecen cuán extremo debería ser un valor en términos de una probabilidad y de la puntuación Z que corresponde a dicha probabilidad. En nuestro ejemplo de la vitamina purificada, los investigadores podrían decidir que si la verosimilitud de un resultado fuera menor del 2% (la probabilidad), rechazarían la hipótesis nula. Estar dentro del 2% inferior de una curva normal significa tener una puntuación Z de aproximadamente –2 ó menor. Por lo tanto, los investigadores establecerían –2 como la puntuación Z de corte en la distribución comparativa para decidir que un resultado es lo suficientemente extremo como para rechazar la hipótesis nula.

Supongamos que los investigadores son todavía más cautelosos en cuanto al rechazo de la hipótesis nula. En ese caso, podrían decidir que rechazarán la hipótesis nula sólo si obtienen un resultado cuyas chances de ocurrir son del 1% ó menos. Entonces podrían calcular la puntuación Z de corte correspondiente al 1%. Conforme a la tabla de áreas bajo la curva normal, para tener un valor dentro del 1% inferior de una curva normal se necesita una puntuación Z de –2,33 ó menor. (*En nuestro ejemplo, una puntuación Z de –2,33 corresponde a 7 meses*). En la figura 6-2 hemos sombreado el 1% de la distribución comparativa, en el que una muestra sería considerada tan extrema que se rechazaría la posibilidad de que surgiera de una distribución como esa.

En general, los investigadores psicológicos utilizan un corte en la distribución comparativa que coincide con una probabilidad del 5% de que un valor sea al menos tan extremo. Es decir, los

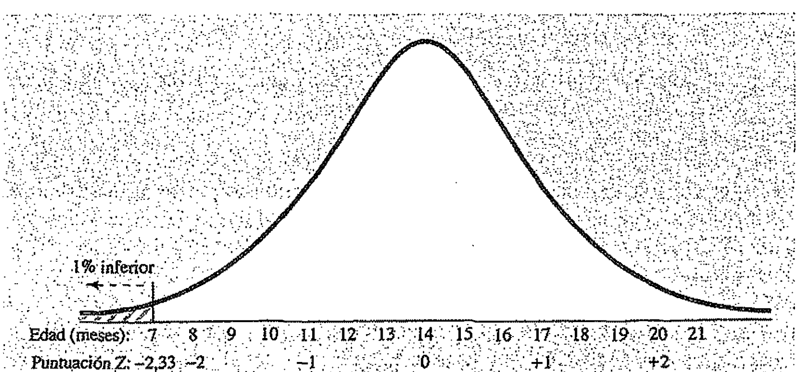


Figura 6-2. Distribución de la edad en la que los bebés comienzan a caminar (datos ficticios).

investigadores rechazan la hipótesis nula si la probabilidad de obtener un resultado tan extremo (si la hipótesis nula fuera verdadera) es menor al 5%. Esta probabilidad generalmente se escribe como "$p<0,05$". No obstante, en algunas áreas de investigación, o cuando los investigadores quieren ser especialmente cautelosos, utilizan un corte del 1% ($p<0,01$).

A estos porcentajes se los denomina **niveles convencionales de significación**. Se describen como nivel de significación 0,05 ó nivel de significación 0,01. Cuando el valor muestral es tan extremo que los investigadores rechazan la hipótesis nula, se dice que el resultado es **estadísticamente significativo**.

4° paso: determinar el valor muestral en la distribución comparativa

El siguiente paso es realizar el estudio y encontrar el resultado real de la muestra. El investigador calcula la puntuación Z correspondiente a la puntuación original de la muestra basándose en la media y el desvío estándar de la distribución comparativa. Esto indica al investigador dónde se ubica su muestra en la distribución comparativa.

Supongamos que los investigadores de nuestro ejemplo realizaron el estudio, y que el bebé que tomó la vitamina especialmente purificada comenzó a caminar a los 6 meses. La media de la distribución comparativa con la que estamos comparando estos resultados es de 14 meses y el desvío estándar de 3 meses. Por lo tanto, un bebé que camina a los 6 meses se ubica 8 meses por debajo de la media, lo que implica un desvío estándar de 2 2/3 por debajo de la media. La puntuación Z correspondiente al bebé de la muestra en la distribución comparativa es –2,67. La figura 6-3 muestra el valor correspondiente al bebé de la muestra en la distribución comparativa.

5° paso: decidir si se rechaza o no la hipótesis nula

Una vez que tenemos claro a) qué puntuación Z debe tener la muestra en la distribución comparativa para poder rechazar la hipótesis nula (paso 3) y b) la puntuación Z real de la muestra (pa-

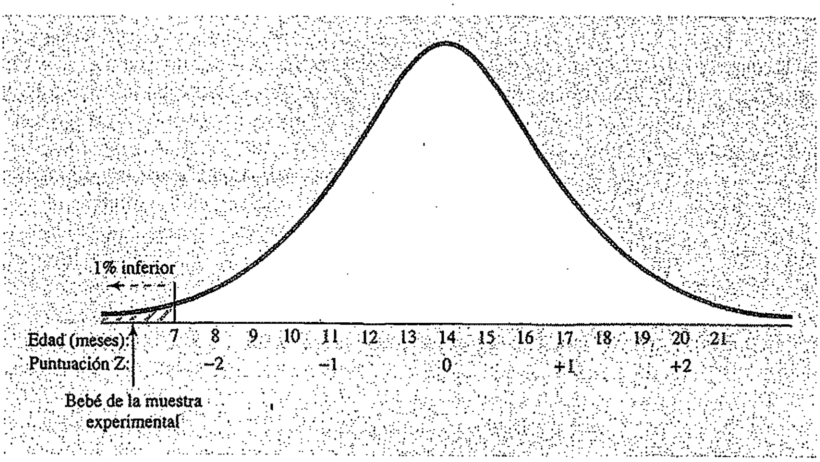


Figura 6-3. Distribución de la edad en que los bebés comienzan a caminar, la cual indica tanto el 1% inferior como la ubicación del bebé que conforma la muestra estudiada (datos ficticios).

so 4), esta decisión es completamente mecánica. Para determinar si se rechaza o no la hipótesis nula, debemos comparar *la puntuación Z necesaria con la puntuación Z real*. Supongamos que en nuestro ejemplo los investigadores determinaron que se rechazaría la hipótesis nula si la puntuación Z de la muestra era menor a –2. Debido a que el resultado real fue –2,67, que es menor a –2, se rechazaría la hipótesis nula. Supongamos que hubieran elegido utilizar el nivel de significación más conservador del 1%. En ese caso, la puntuación Z necesaria hubiera sido –2,33. Siendo la puntuación Z real igual a –2,67, aun con este criterio más conservador se rechazaría la hipótesis nula.

Si los investigadores rechazan la hipótesis nula, lo que queda es la hipótesis de investigación. En este ejemplo, los investigadores pueden inferir que los resultados de su estudio sostienen la hipótesis de investigación que indica que los bebés que toman la vitamina especialmente purificada comienzan a caminar antes que los otros bebés.

¿Qué implica rechazar o no la hipótesis nula?

Queremos hacer hincapié en dos puntos relacionados con el tipo de conclusiones que podemos sacar a partir del proceso de prueba de hipótesis. En primer lugar, supongamos que rechazamos la hipótesis nula y que los resultados sostienen la hipótesis de investigación (como en nuestro ejemplo de la vitamina). Los investigadores aún no dirían que el resultado "prueba" la hipótesis de investigación o que los resultados muestran que la hipótesis es "verdadera". Esas palabras son demasiado fuertes en este caso, ya que las conclusiones a las que se llega a través de estudios de investigación siempre se basan en probabilidades. En la prueba de hipótesis, se basan en la poca probabilidad de obtener determinado resultado si la hipótesis nula fuera verdadera. Decir que las conclusiones están **comprobadas** o que son **verdaderas** sería una exageración. Tales afirmaciones son correctas para la lógica o la matemática, pero utilizar estas palabras con respecto a conclusiones resultantes de una investigación científica es completamente poco profesional. (Es correcto usar la palabra "verdadero" cuando se habla hipotéticamente, por ejemplo, "si la hipótesis fuera verdadera, entonces...", pero no al hablar de una conclusión real).

En segundo lugar, y tal como lo mencionamos anteriormente, cuando un resultado no es lo suficientemente extremo como para que rechacemos la hipótesis nula, no decimos que el resultado "sostiene la hipótesis nula". Un resultado que no es lo suficientemente determinante como para que rechacemos la hipótesis nula sólo implica que el estudio no fue concluyente. Los resultados pueden no ser lo suficientemente extremos como para rechazar la hipótesis nula, pero la hipótesis nula podría ser falsa (y la hipótesis de investigación verdadera). Supongamos que en nuestro ejemplo la vitamina especialmente purificada tuviera sólo un efecto leve, pero aún así real. En ese caso, no esperaríamos que ningún bebé que hubiera tomado la vitamina purificada camine mucho antes que los otros bebés; por lo tanto, no podríamos rechazar la hipótesis nula aunque esta fuera falsa.

La cuestión es que demostrar que la hipótesis nula es verdadera implicaría demostrar que realmente no existe diferencia entre las poblaciones. No obstante, siempre es posible que esa diferencia exista pero que sea mucho menor de lo que el estudio en particular podría detectar. Por lo tanto, cuando un resultado no es lo suficientemente extremo como para rechazar la hipótesis nula, los investigadores por lo general sólo dicen que los resultados no son concluyentes. Sin embargo, algunas veces, si los estudios se han realizado utilizando grandes cantidades de procedimientos con un nivel de medición muy precisa, la evidencia puede crear fundamentos en cuanto a la exactitud aproximada de determinada hipótesis nula. Además, algunas veces, los investigadores hablando informalmente describen la imposibilidad de rechazar una hipótesis nula como un resultado que "sostiene la hipótesis nula". Sin embargo, técnicamente, la expresión anterior es, por lo general, demasiado fuerte. (Más adelante, en este capítulo y en el capítulo 8, continuaremos tratando este tema).

Resumen de los pasos de la prueba de hipótesis

A continuación presentamos un resumen de los cinco pasos de la prueba de hipótesis:

1. Replantear el problema en función de la hipótesis de investigación e hipótesis nula de las poblaciones.
2. Determinar las características de la distribución comparativa.
3. Determinar el punto muestral de corte en la distribución comparativa, a partir del cual debería rechazarse la hipótesis nula.
4. Determinar el valor muestral en la distribución comparativa.
5. Comparar los valores de los pasos 3 y 4 para decidir si se rechaza o no la hipótesis nula.

Otro ejemplo de prueba de hipótesis

Aquí presentamos otro ejemplo ficticio. Dos psicólogos especializados en personalidades despreocupadas están analizando la teoría de que la felicidad surge de experiencias positivas. En particular, los investigadores sostienen que si a una persona le sucede algo muy afortunado se pondrá muy feliz y continuará estándolo durante mucho tiempo. Por lo tanto, planifican el siguiente experimento: de todas las personas adultas norteamericanas se seleccionará una al azar y se le regalará 1 millón de dólares, y seis meses después se medirá la felicidad de esa persona. En este ejemplo ficticio ya se conoce cuál es la distribución correspondiente a la felicidad para la población general de adultos de Norteamérica, que es la que representa la figura 6-4. En la prueba que se utiliza, la felicidad tiene un valor medio de 70, el desvío estándar es 10, y la distribución es aproximadamente normal.

Los psicólogos siguen el mismo procedimiento de prueba de hipótesis utilizado en el ejemplo de la vitamina purificada. Consideran cuál es el nivel de felicidad que debería sentir la persona analizada para rechazar con confianza la hipótesis nula (que implica que recibir esa cantidad de dinero no hace que las personas se sientan más felices 6 meses después). Si el resultado obtenido por los investigadores muestra un muy alto nivel de felicidad, los psicólogos rechazarán la hipótesis nula y concluirán que obtener 1 millón de dólares probablemente hace sentir más felices a las personas 6 meses después del hecho. Pero si el resultado no lo es suficientemente extremo, los in-

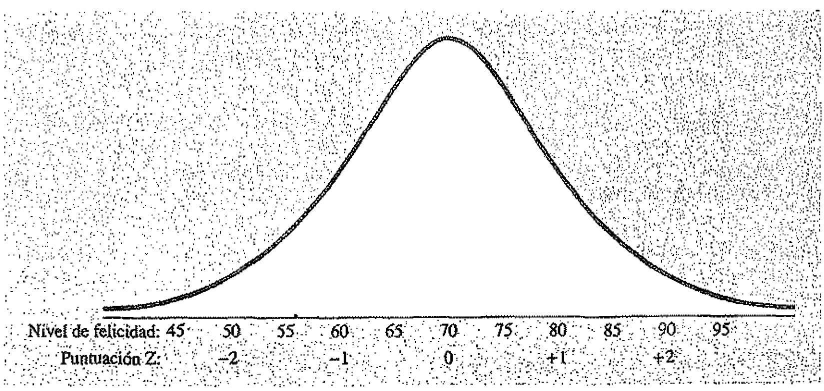


Figura 6-4. Distribución de registros del nivel de felicidad (datos ficticios).

Cuadro 6-1.
Ser o no ser, pero ¿es posible no ser? Cuándo sí debe aceptarse la hipótesis nula.

La hipótesis nula establece que no existe diferencia entre las poblaciones representadas por diferentes grupos o condiciones experimentales. Como hemos observado, la regla general en estadística es que un estudio no puede determinar que la hipótesis nula sea verdadera. Un estudio sólo puede indicar que no es posible rechazar la hipótesis nula, es decir, que ese estudio simplemente no aporta información. Obviamente, esos estudios no suelen publicarse, aunque de hecho se podría evitar una gran cantidad de trabajo si las personas supieran que ciertas intervenciones, medidas o experimentos no funcionaron anteriormente. En realidad, Greenwald (1979) informa que en algunas ocasiones se ha considerado por mucho tiempo que algunas ideas eran verdaderas sólo porque unos pocos estudios así lo mostraban, mientras que muchos otros, no publicados, habían mostrado lo contrario.

> Frick (1995) ha señalado un problema aún más serio con respecto al inflexible desinterés por la hipótesis nula: Algunas veces puede ser verdad que determinado elemento no tenga ningún efecto sobre otro. Este hecho no significa que exista una relación cero, que no exista diferencia en absoluto (la ausencia absoluta de diferencia es un resultado improbable en la mayoría de los casos). Sólo significaría que el efecto de un elemento sobre el otro es tan pequeño que no tiene ninguna importancia práctica o teórica.

El problema es saber cuándo inferir que la hipótesis nula (o algo cercano a ella) podría ser verdadera. Frick (1995) propone tres criterios. Primero, la hipótesis nula debería parecer posible. Segundo, obviamente los resultados del estudio deberían ser coherentes con la hipótesis nula, por lo cual no debería existir ninguna otra forma evidente de interpretarlos. Tercero, y más importante aún, el investigador debe haber realizado un gran esfuerzo por descubrir el efecto cuya inexistencia pretende inferir. Entre otras cosas, esto implica analizar una gran muestra y emplear una medición sumamente completa y susceptible. Si el estudio es un experimento, es importante que se haya intentado producir el efecto utilizando una fuerte manipulación y rigurosas condiciones de prueba.

Frick señala que todo esto deja un elemento subjetivo en cuanto a la aceptación de la hipótesis nula: ¿Quién decide si el esfuerzo del investigador fue lo suficientemente importante? Pero nos guste o no, las decisiones subjetivas son parte de la ciencia. Por ejemplo, los editores deben decidir si un tema es lo suficientemente importante como para brindarle el espacio en su revista. Más aún, a pesar de todo, la hipótesis nula es aceptada en muchas ocasiones (por ejemplo, muchos psicólogos aceptan la hipótesis nula con respecto al efecto de la ESP-*Extrasensory Perception*, Percepción extrasensorial). Es más conveniente debatir los fundamentos para la aceptación de la hipótesis nula que, simplemente, aceptarla.

¿Cuál es el objetivo de toda esta argumentación? Queda claro que no rechazar la hipótesis nula no es lo mismo que sostenerla. Pero Frick nos recuerda que existen situaciones en las que la evidencia debería convencernos de que algo similar a la hipótesis nula podría ser la situación verdadera.

vestigadores concluirán que no existe suficiente evidencia para rechazar la hipótesis nula y los resultados del experimento serían, entonces, no concluyentes.

Ahora analicemos el procedimiento de prueba de hipótesis más detalladamente según este ejemplo, siguiendo los pasos resumidos anteriormente.

1. **Replantear el problema en función de hipótesis de investigación e hipótesis nula de las poblaciones.** Las poblaciones de interés son las siguientes:

Población 1: personas que hace 6 meses recibieron 1 millón de dólares.
Población 2: personas que hace 6 meses no recibieron 1 millón de dólares.

La predicción de los psicólogos especializados en personalidad, basándose en esta teoría de la felicidad, es que las personas que forman la población 1 se sentirán en general más felices que las personas que forman la población 2 (en símbolos, $M_1 > M_2$). La hipótesis nula implica que las personas que forman la población 1 (los que recibieron 1 millón de dólares) no se sentirán más felices que las personas que forman la población 2 (aquellos que no recibieron 1 millón de dólares).

2. **Determinar las características de la distribución comparativa.** Llegado el momento querremos comparar nuestra observación con la situación que se presentaría si la hipótesis nula fuera verdadera (para comprobar si podemos rechazar ese escenario). Si la hipótesis nula es verdadera, la distribución de las poblaciones 1 y 2 serán iguales. Sabemos cuál es la distribución de la población 2, así que puede servirnos como distribución comparativa.

3. **Determinar el punto muestral de corte en la distribución comparativa, a partir del cual debería rechazarse la hipótesis nula.** ¿Qué tipo de observación sería suficientemente convincente como para rechazar la hipótesis nula? En este caso, supongamos que los investigadores decidieron por adelantado rechazar la hipótesis nula, por ser demasiado improbable, si los resultados pudieran ocurrir menos de un 5% de las veces si esa hipótesis nula fuera verdadera. Debido a que sabemos que la distribución comparativa es normal, podemos determinar a partir de la tabla de áreas bajo la curva normal que el 5% superior de los valores comienzan en una puntuación Z de aproximadamente 1,64. (Siendo la media de la distribución comparativa igual a 70 y el desvío estándar igual a 10, la hipótesis nula sería rechazada si el resultado de la muestra fuera igual o mayor a 86,4. Es decir, siguiendo el método usual para convertir una puntuación Z en una puntuación original, 1,64 x 10 = 16,4, lo que sumado a la media de 70 da 86,4).

4. **Determinar el valor muestral en la distribución comparativa.** Observemos ahora los resultados: seis meses después de entregar 1 millón de dólares a la persona elegida al azar, los investigadores entregan a su ahora adinerado participante la prueba de nivel de felicidad. La puntuación de la persona es 80. Como puede verse en la figura 6-4, una puntuación de 80 corresponde a una puntuación Z de +1 en la distribución comparativa.

5. **Comparar los registros de los pasos 3 y 4 para decidir si se rechaza o no la hipótesis nula.** La puntuación Z mínima, necesaria para rechazar la hipótesis nula, ha sido establecida en +1,64 (la puntuación Z que corresponde al 5% del nivel de significación), y la puntuación Z correspondiente al individuo de la muestra es sólo +1. Por lo tanto, la muestra no es lo suficientemente extrema como para darnos fundamentos para rechazar la hipótesis nula. La hipótesis nula no puede rechazarse, y los resultados del experimento no son concluyentes. Los investigadores describen tal resultado como "no significativo estadísticamente". La figura 6-5 muestra la distribución comparativa con el 5% superior sombreado y la ubicación del millonario que conforma la muestra.

Un dato interesante es que Brickman, Coates y Janoff-Bulman (1978) realizaron un estudio más elaborado basándose en la misma cuestión, analizando a ganadores de la lotería como ejemplos de personas a las que les ocurrían hechos repentinos muy positivos. Sus resultados fueron similares a los de nuestro ejemplo ficticio: 6 meses después, el grupo ganador de dinero no era mucho más feliz que las personas que no habían ganado ese dinero. Además, descubrieron que otro grupo estudiado por ellos, personas que habían quedado parapléjicas a causa de accidentes, 6 meses después no eran mucho menos felices que otras personas. Estos investigadores analizaron cantidades bastante grandes de individuos e investigaron el tema de diversas maneras. Su conclusión fue que si un hecho importante en verdad provoca algún efecto duradero relacionado con la felicidad, el efecto probablemente no es muy grande. Aparentemente, ganarse la lotería no es la respuesta. (En otros estudios, p. ej. Suh, Diener & Fijita, 1996, se ha encontrado el mismo patrón).

PRUEBAS DE HIPÓTESIS DE UNA Y DOS COLAS

Hasta aquí nuestros ejemplos de prueba de hipótesis se han basado en situaciones en las que nos interesaba sólo una dirección del resultado. En el ejemplo de la vitamina purificada, los investigadores estaban interesados en saber si el bebé caminaría **antes** que otros bebés. En el ejemplo sobre la felicidad, los psicólogos especializados en el estudio de la personalidad esperaban que la persona que recibiera 1 millón de dólares fuera **más feliz** que las otras. Los investigadores que realizaron estos estudios no estaban realmente interesados en la posibilidad de que el suministro de las vitaminas especialmente purificadas pudiera causar que los bebés tardaran más en comenzar a caminar, o que la persona que recibió 1 millón de dólares pudiera en realidad ser menos feliz.

Hipótesis direccional y pruebas de una cola

Los estudios acerca de la vitamina purificada y la felicidad son ejemplos que involucran **hipótesis direccionales.** En cada caso, los investigadores estaban interesados en una dirección específica del efecto. Es importante observar que cuando un investigador propone una hipótesis direccional, la hipótesis nula correspondiente es, también, en cierto sentido direccional. Si la hipótesis de investigación establece *que obtener 1 millón de dólares hará más feliz a una persona, la hipótesis nula es-*

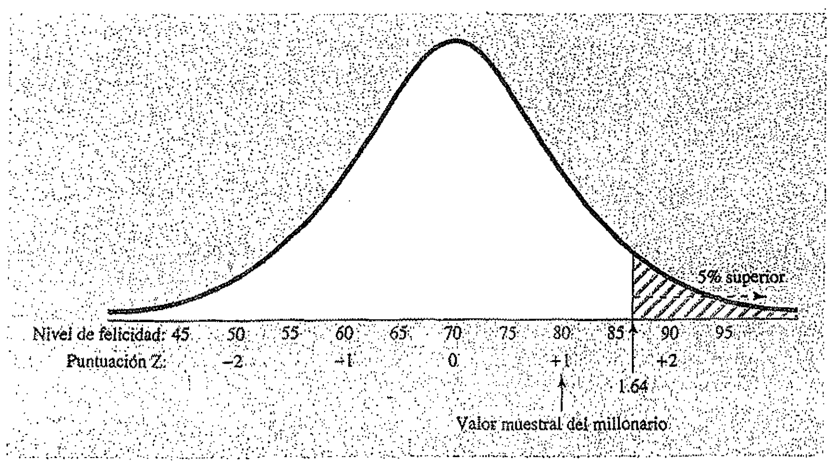


Figura 6-5. Distribución de valores del nivel de felicidad con el 5% superior sombreado y la ubicación del millonario que conforma la muestra (datos ficticios).

tablece que el dinero no producirá ningún efecto o hará menos feliz a esa persona. (Expresado en símbolos, si la hipótesis de investigación es $\mu_1 > \mu_2$, entonces la hipótesis nula será $\mu_1 \leq \mu_2$). Por lo tanto, como ya hemos observado, en la figura 6-5, por ejemplo, para rechazar la hipótesis nula la muestra debía arrojar un valor que se ubicara dentro del 5% superior, el extremo o cola superior de la distribución comparativa. (A los fines de rechazar la hipótesis nula, un valor ubicado en la otra cola sería considerado del mismo modo que un valor ubicado en el medio de la distribución). Por esta razón, la prueba de una hipótesis direccional se denomina **prueba de una cola**.

Hipótesis no direccional y pruebas de dos colas

Sin embargo, a veces una hipótesis de investigación implica simplemente que una población será diferente de la otra, sin especificar si la diferencia la marcarán valores más altos o más bajos. Por ejemplo, un psicólogo especializado en organizaciones empresariales puede estar interesado en el impacto provocado en la productividad por un programa de capacitación en relaciones sociales. Es posible que el programa mejore la productividad al hacer más placentero el ambiente de trabajo. Pero también es posible que perjudique la productividad por incentivar a las personas a que practiquen relaciones sociales en lugar de trabajar. En este caso, la hipótesis de investigación implicaría que el programa de relaciones sociales cambie el nivel de productividad. La hipótesis nula implicaría que el programa no afecte la productividad en ningún sentido. Es decir, expresado en símbolos, la hipótesis de investigación sería $\mu_1 \neq \mu_2$, y la hipótesis nula sería $\mu_1 = \mu_2$.

Siempre que una hipótesis de investigación establezca una diferencia, sin indicar la dirección de esa diferencia, se la denomina **hipótesis no direccional**. Para probar la significación de una hipótesis no direccional, uno debe analizar si un valor es extremo en cualquiera de las dos colas de la distribución comparativa. Por lo tanto, a esta prueba se la denomina **prueba de dos colas**.

Determinación de puntos de corte en pruebas de dos colas

Las pruebas de dos colas presentan una complicación especial. Supongamos que el investigador selecciona un nivel de significación del 5%. En una prueba de una cola, el investigador rechaza la hipótesis nula si la oservación muestral se ubica dentro de uno de los extremos que contiene el 5% de la distribución comparativa. En una prueba de dos colas, podría suponerse que el investigador utilizaría el 5% superior cuando el valor es extremo en dirección hacia arriba, y el 5% inferior cuando el valor es extremo en dirección hacia abajo. Sin embargo, si el investigador hiciera esto, existiría un total del 10% de la distribución comparativa dentro del cual la hipótesis nula podría ser rechazada. El nivel de significación en realidad sería del 10%, porcentaje que la mayoría de los investigadores consideraría muy peligroso. (Es decir, con un 10% de nivel de significación, uno podría rechazar la hipótesis nula con mucha facilidad aun cuando ésta fuera verdadera).

Existe una solución para este problema. Al realizar una prueba de dos colas, se divide el porcentaje de significación entre las dos colas. Con un nivel de significación del 5%, se rechazaría la hipótesis nula sólo si la muestra fuera tan extrema que se ubicara dentro del 2 1/2 % superior o dentro del 2 1/2 % inferior. De este modo, la posibilidad total de que la hipótesis nula sea verdadera, determinada con anterioridad a la realización del estudio, se mantiene en un total del 5%.

Es importante señalar que al utilizar una prueba de dos colas, las puntuaciones Z de corte para un nivel del 5% son +1,96 y –1,96. En el caso de una prueba de una cola, el corte no era tan extremo, +1,64 y –1,64, pero sólo se tenía en cuenta un lado de la distribución. La figura 6-6a representa esas situaciones. Utilizando un nivel de significación del 1%, una prueba de dos colas (0,5% en cada cola) presenta cortes de +2,58 y –2,58, mientras que los cortes en una prueba de una cola serían de +2,33 ó –2,33 (*véase* figura 6-6b).

¿Cuándo utilizar pruebas de una o dos colas?

Resulta más fácil rechazar la hipótesis nula con una prueba de una cola que con una prueba de dos colas, ya que el valor de la muestra no necesita ser tan extremo para que el resultado experimental sea significativo. Sin embargo, esto tiene su costo, ya que con las pruebas de una cola, si el resultado es extremo en la dirección opuesta a la esperada, no puede considerarse significativo y no importa cuán extremo haya sido ese resultado.

En principio, se planifica una prueba de una cola cuando se trabaja con una hipótesis claramente direccional, y de dos colas cuando se trabaja con una hipótesis claramente no direccional. En la práctica, la decisión no resulta tan simple. Incluso cuando una teoría predice claramente un resultado determinado, a veces descubrimos que el resultado es justamente el opuesto de lo que esperábamos, y en ocasiones ese resultado opuesto puede ser realmente más interesante. (¿Qué hubiera sucedido si, como ocurre en todos los cuentos de hadas sobre genios y peces que conceden deseos, recibir 1 millón de dólares y cumplir casi todos sus deseos hubiera hecho de ese individuo una persona infeliz? El resultado hubiera sido realmente muy interesante). Utilizando las pruebas de una cola corremos el riesgo de tener que ignorar resultados posiblemente importantes.

Debido a estas consideraciones, la utilización de las pruebas de una cola es discutida, aun cuando la hipótesis sea claramente direccional. Para mayor seguridad, muchos investigadores utilizan pruebas de dos colas tanto para hipótesis direccionales como no direccionales. Si el resultado de la prueba de dos colas es significativo, entonces el investigador analiza el patrón de los datos hallados para determinar la dirección del resultado, considerando al estudio significativo en esa dirección.[2] Cabe mencionar que, en la práctica, este es un procedimiento conservador, por el hecho de que siendo los puntos de corte más extremos para una prueba de dos colas, es menos verosímil que una prueba de dos colas dé un resultado significativo. Por lo tanto, si se obtiene un resultado significativo con una prueba de dos colas, uno puede estar más seguro de sus conclusiones. De hecho, en la mayoría de las publicaciones cientifcas psicológicas, a menos que el investigador indique específicamente que utilizó una prueba de una cola, en líneas generales se supone que utilizó una prueba de dos colas.

No obstante, cabe recordar que, por lo general, la conclusión final no es afectada realmente por el hecho de que el investigador utilice una prueba de una o dos colas. Según nuestra experiencia, usualmente los resultados de las investigaciones o son tan extremos que serían considerados significativos a través de cualquier estándar razonable, o están tan lejos de serlo que no serían considerados significativos a través de ningún procedimiento.

¿Qué sucede cuando un resultado arroja conclusiones menos precisas? La decisión del investigador en cuanto a las pruebas de una o dos colas adquiere mayor importancia. En ese caso, el investigador intentará utilizar el método que arroje la conclusión más exacta y menos controvertida, ya que la idea es dejar que hasta donde sea posible, la naturaleza, y no la decisión del investigador, determine la conclusión. Más aún, cuando un resultado no es completamente claro en uno u otro sentido, la mayoría de los investigadores se sentirían incómodos al sacar conclusiones definitivas sin realizar otros estudios.

[2] Leventhal y Huynh (1996) sostienen que este procedimiento en realidad es incorrecto. Si uno está probando una hipótesis no direccional, sólo debería sacar conclusiones no direccionales. Sugieren que un mejor procedimiento sería utilizar una "prueba direccional de dos colas", que en realidad son dos pruebas simultáneas de una cola (una en cada dirección). Así, si un investigador quisiera establecer un nivel de significación total de 0,05, utilizaría una prueba direccional de dos colas, en la que cada una de las dos subdivisiones de una cola utilizaría el nivel 0,025. En cuanto a decidir si un resultado es significativo o no, el método de Leventhal y Huynh produce un resultado idéntico al de la prueba de uso más común, no direccional de dos colas. El razonamiento de Leventhal y Huynh sobre las pruebas de dos colas parece más lógico (además de tener otras ventajas técnicas). Sin embargo, debido a que los investigadores aún no han adoptado ese método (y dado que el resultado es el mismo), en este libro utilizamos el método más tradicional.

Ejemplo de prueba de hipótesis utilizando una prueba de dos colas

Aquí presentamos otro ejemplo ficticio, pero esta vez utilizando una prueba de dos colas. Un grupo de psicólogos clínicos de un centro residencial de tratamiento psiquiátrico creen haber desarrollado un nuevo tipo de terapia que aliviará, en mayor grado que la terapia que se está utilizando en ese momento, la depresión de los pacientes. Sin embargo, como sucede con cualquier tratamiento, no se puede descartar la posibilidad de que provoque peores resultados en algún paciente. Por lo tanto, los investigadores probarán una hipótesis no direccional.

Los psicólogos procederán de la siguiente manera: seleccionarán al azar un paciente que recién ingrese para suministrarle la nueva terapia en lugar de la usual. (Por supuesto que en un estu-

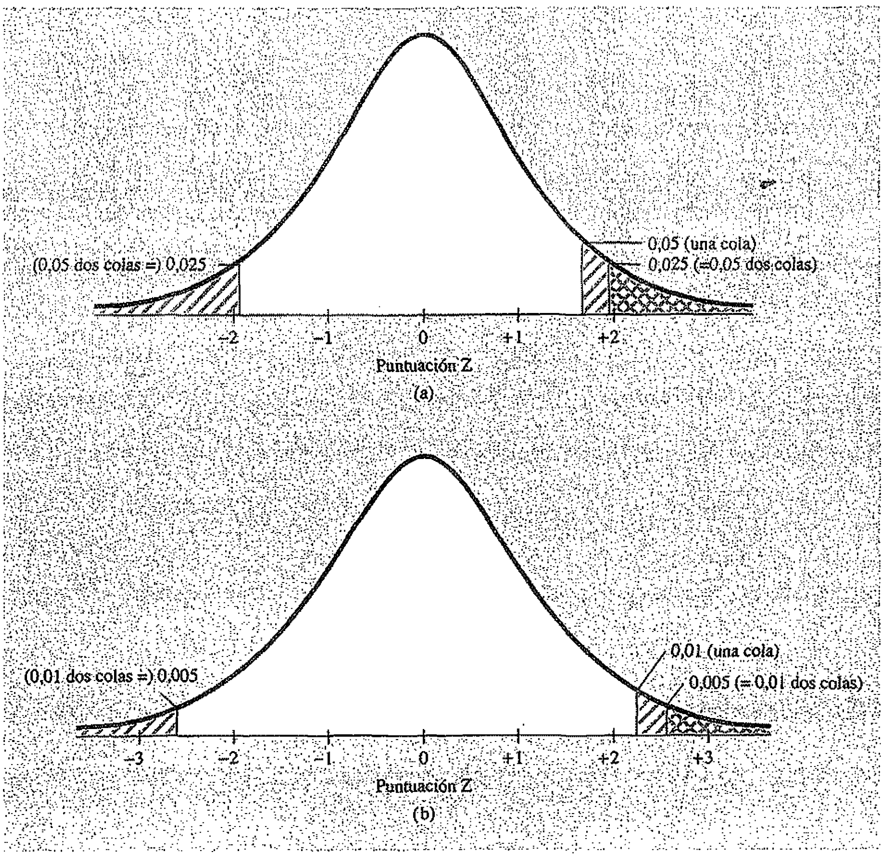


Figura 6-6. *Comparación de puntos de corte según el nivel de significación para pruebas de una y dos colas:* (a) nivel de significación 0,05; (b) nivel de significación 0,01. (Las pruebas de una cola en estos ejemplos suponen que se predecía un valor alto).

dio real se seleccionaría más de un paciente, pero supongamos que una sola persona ha sido capacitada para realizar la nueva terapia y que tiene tiempo para tratar sólo a un paciente). La depresión del paciente se medirá con una escala de depresión estándar que se aplica automáticamente a todos los pacientes después de 4 semanas. Esa escala ha sido aplicada a los pacientes durante un largo tiempo en este centro de tratamiento. Por lo tanto, es posible determinar por adelantado "en aquellos pacientes que recibieron la terapia usual" la distribución de los valores del nivel de depresión a las 4 semanas. En nuestro ejemplo ficticio, esa distribución sigue una curva normal con una media de 69,5 y un desvío estándar de 14,1. (Las cifras mencionadas se aproximan a los valores de depresión obtenidos en una encuesta nacional de 75.000 pacientes psiquiátricos a los que se les suministró el MMPI, una prueba estándar ampliamente utilizada; Dahlstrom, Larbar, & Dahlstrom, 1986). La figura 6-7 muestra esta distribución.

El procedimiento de prueba de hipótesis se realiza, entonces, de la siguiente manera:

1. **Replantear el problema en función de hipótesis de investigación e hipótesis nula de las poblaciones.** Las dos poblaciones de interés son:

Población 1: pacientes con diagnóstico de depresión que recibieron la nueva terapia.
Población 2: pacientes con diagnóstico de depresión que recibieron la terapia estándar.

La hipótesis de investigación supone que, al medir la depresión 4 semanas después del ingreso, los pacientes que reciben la nueva terapia (población 1) tendrán un valor diferente al de los pacientes que reciben la terapia actual (población 2). En símbolos, la hipótesis de investigación es $M_1 \neq M_2$. Lo contrario a la hipótesis de investigación, la hipótesis nula, supone que los pacientes que reciben la nueva terapia tendrán el mismo nivel de depresión que los pacientes que reciben la terapia usual. (Es decir, el nivel de depresión medido después de 4 semanas será el mismo para la población 1 y 2). En símbolos, la hipótesis nula es: $\mu_1 = \mu_2$.

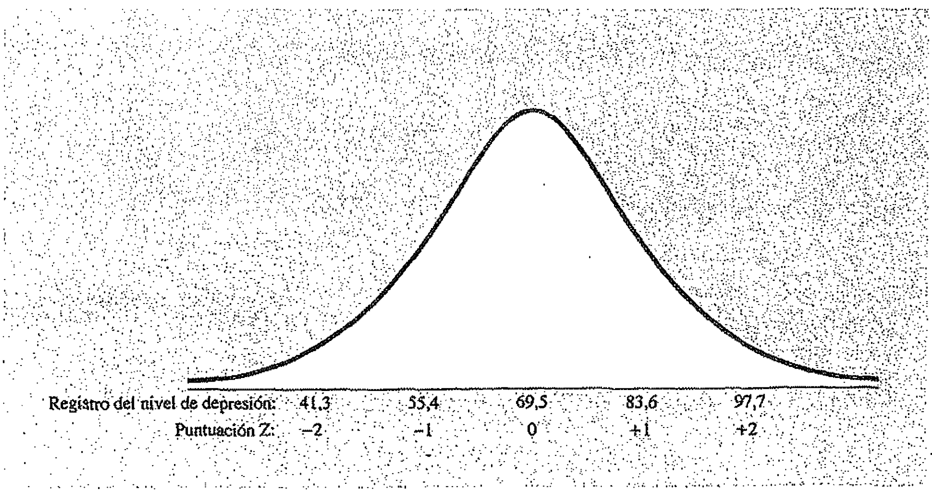


Figura 6-7. Distribución de los valores de la escala de depresión MMPI a 4 semanas del ingreso, correspondientes a pacientes psiquiátricos a los que se les diagnosticó depresión y que reciben la terapia estándar (datos ficticios).

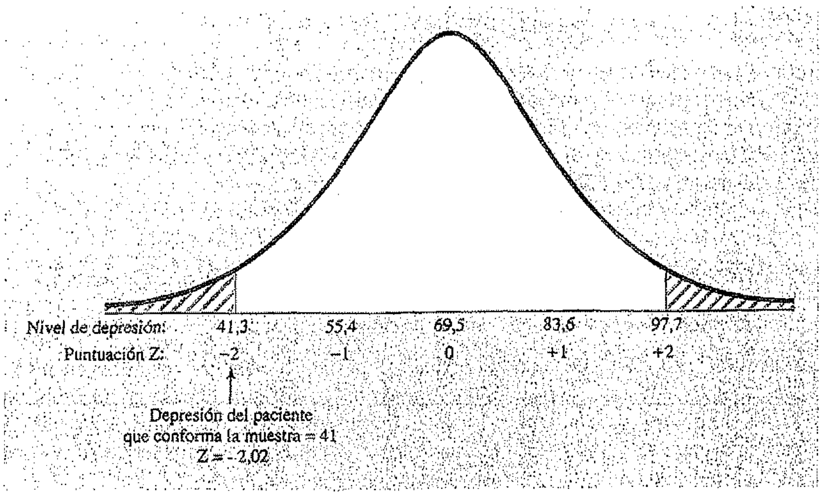


Figura 6-8. Distribución de los valores de la escala de depresión MMPI con el 2 1/2 % superior e inferior sombreado, el cual indica la ubicación del paciente que conforma la muestra y que recibió la nueva terapia. (datos ficticios).

2. **Determinar las características de la distribución comparativa.** Si la hipótesis nula es verdadera, las distribuciones de las poblaciones 1 y 2 serán iguales. Conocemos la distribución de la población 2, por lo tanto, puede servir como distribución comparativa. Como ya dijimos, presenta una curva normal con $\mu = 69{,}5$ y $\sigma = 14{,}1$.

3. **Determinar el punto muestral de corte en la distribución comparativa, a partir del cual debería rechazarse la hipótesis nula.** El equipo de psicólogos clínicos selecciona un nivel de significación del 5%. Los investigadores han preparado una hipótesis no direccional, por lo que se debe utilizar una prueba de dos colas. Esto significa que la hipótesis nula será rechazada sólo si el valor del nivel de depresión del paciente en la distribución comparativa se encuentra dentro del 2 1/2 % superior o inferior de esa distribución. Expresados en puntuaciones Z, los puntos críticos son +1,96 y −1,96 (*véase* figura 6-8).

4. **Determinar el valor muestral en la distribución comparativa.** El paciente que recibió la nueva terapia fue medido 4 semanas después de su ingreso. El valor del paciente en la escala de depresión fue de 41, lo que es igual a una puntuación Z de −2,02 en la distribución comparativa.

5. **Comparar los valores de los pasos 3 y 4 para decidir si se rechaza o no la hipótesis nula.** Una puntuación Z de −2,02 se ubica apenas por debajo de la puntuación Z de −1,96, que es donde comienza el 2 1/2 % inferior de la distribución comparativa. Se trata de un resultado tan extremo que es improbable que hubiera ocurrido si el paciente representara una población que no fuera diferente de la población 2. Por lo tanto, los psicólogos clínicos rechazaron la hipótesis nula. El resultado sostiene la hipótesis de investigación que implica que la nueva terapia realmente produce cambios en el nivel de depresión de los pacientes.

CONTROVERSIAS Y LIMITACIONES

En los últimos años ha surgido una controversia importante acerca de la propia prueba de significación, con un movimiento organizado por parte de un pequeño pero resonante grupo de psicólogos que pretendían prohibirlas. La sugerencia es radical, y sus consecuencias tendrían un enorme alcance (durante al menos medio siglo casi todas las investigaciones psicológicas han utilizado las pruebas de significación). Probablemente, en las más importantes revistas científicas especializadas en psicología se haya escrito recientemente mucho más acerca de esta controversia. El comienzo de una publicación reciente ilustra la fuerza que ha adquirido el debate:

> No es verdad que un grupo de activistas radicales tomaran a 10 estadísticos y 6 editores como rehenes en la Convención de la Sociedad Americana de Psicología de 1996 y corearan "apoyen la prohibición total de las pruebas" y "anulen la (hipótesis) nula". (Abelson, 1997, p.12).

Dado que se trata, hasta ahora, de la más enérgica e importante controversia surgida en años con respecto a la estadística aplicada a la psicología, trataremos los distintos temas relacionados con ella al menos en tres diferentes oportunidades. En este capítulo nos concentraremos en algunos desafíos básicos para la prueba de hipótesis. En los capítulos 7 y 8 tocaremos otros temas relacionados con aspectos de la prueba de hipótesis que enseñaremos en esos capítulos.

Antes de exponer esta controversia, queremos asegurar al alumno que no está aprendiendo la prueba de hipótesis inútilmente. No importa lo que suceda en el futuro, ya que es absolutamente necesario comprender la prueba de hipótesis para poder encontrar el sentido de todas las publicaciones científicas publicadas en el pasado. Más aún, a pesar de la vehemente controversia que ha surgido en los últimos años, es sumamente extraño ver nuevas publicaciones que no utilicen la prueba de significación, por lo que resulta dudoso que ocurra algún cambio importante en un futuro cercano. Finalmente, aun si se abandonara por completo la prueba de hipótesis, las alternativas (que involucran procedimientos que enseñaremos en los capítulos 7 y 8) requieren la comprensión de prácticamente toda la lógica y de todos los procedimientos que tratamos aquí.

¿Cuál es entonces la gran controversia? Algunos puntos del debate están relacionados con sutiles temas de lógica. Por ejemplo, una postura plantea si tiene sentido preocuparse por rechazar la hipótesis nula cuando es extremadamente improbable que resulte verdadera una hipótesis que supone que no se produce ningún tipo de efecto. Tratamos este tema brevemente en el cuadro 6-1.

Otro de los temas está relacionado con los fundamentos de la prueba de hipótesis en relación con las poblaciones y las muestras, debido a que en la mayoría de los experimentos las muestras que utilizamos de la población definible no son seleccionadas de manera aleatoria. En el capítulo 5 tratamos algunos puntos relacionados con este tema. Finalmente, algunos han cuestionado lo adecuado de llegar a la conclusión de que si la información es inconsistente con la hipótesis nula, esto debe ser considerado como evidencia de la hipótesis de investigación. Esta controversia es bastante técnica, pero nuestra propia opinión es que lo que estamos haciendo es razonable, conforme a recientes consideraciones sobre estos temas, (*véase*, p. ej. Cortina & Dunlop, 1997).

De todos modos, la queja más considerada contra las pruebas de significación, y que ha obtenido el acuerdo prácticamente universal, es que las pruebas están mal utilizadas. De hecho, los opositores de las pruebas de significación sostienen que aun si no existieran otros inconvenientes con respecto a las pruebas, éstas deberían ser prohibidas, simplemente por ser utilizadas con tanta frecuencia de un modo tan inadecuado. Son dos los casos de pruebas que se utilizan inadecuadamente. Una podemos analizarla ahora, la otra deberá esperar hasta que hayamos tratado un tema que enseñaremos en el capítulo 8.

Uno de los principales usos inapropiados de las pruebas es la tendencia de los investigadores a decidir que, si un resultado no es significativo, queda demostrado que la hipótesis nula es verdadera. Repetidamente hemos subrayado que cuando no se rechaza la hipótesis nula, los resultados no son concluyentes. El error de llegar a la conclusión de que la hipótesis nula es verdadera, debido a la imposibilidad de rechazarla, es extremadamente serio, ya que pueden considerarse falsos importantes métodos y teorías sólo porque determinado estudio no logró resultados lo suficientemente fuertes. (Como veremos en el capítulo 8, es bastante fácil que una hipótesis de investigación verdadera no resulte significativa sólo porque el estudio se realizó con pocas personas o porque las medidas no eran muy precisas. De hecho, Hunter (1997) sostiene que en aproximadamente el 60% de los estudios psicológicos es probable que obtengamos resultados no significativos aun cuando la hipótesis de investigación sea realmente verdadera).

¿Cuál es entonces la solución? El consenso general parece determinar que deberíamos mantener las pruebas de significación, pero preparando mejor a nuestros alumnos para que no las utilicen de manera inadecuada (a esto se debe que se haya hecho tanto hincapié en esos temas a lo largo del libro), es decir que deberíamos cuidarnos de no perder una herramienta valiosa sólo porque no se la utilice en manera adecuada. Con el fin de tratar esta controversia, la APA estableció un comité formado por eminentes psicólogos renombrados por su experiencia en estadística. En el informe provisorio del Cuerpo de trabajo sobre inferencia estadística [*Task Force on Statistical Inference*] de la APA (1996), llegaron a la siguiente conclusión:

> Respaldamos una política de inclusión que admita en el arsenal del científico de investigación cualquier procedimiento que **apropiadamente** arroje algo de luz sobre el fenómeno de interés. En este sentido, el Cuerpo de Trabajo no respalda ninguna acción que pueda ser interpretada como prohibición del uso de la prueba de significación de la hipótesis nula o de los valores *p* en investigaciones y publicaciones psicológicas. (p. 2)

LA PRUEBA DE HIPÓTESIS SEGÚN SE DESCRIBE EN LAS PUBLICACIONES CIENTÍFICAS

En líneas generales, las pruebas de hipótesis aparecen en las publicaciones científicas como parte de uno de los procedimientos estadísticos específicos que enseñaremos en capítulos posteriores. Para cada resultado de interés el investigador usualmente indica primero si el resultado fue "estadísticamente significativo". Luego, por lo general el investigador da el nombre de la técnica específica utilizada para determinar las probabilidades, como puede ser una prueba t, ó F, ó χ^2 (tratadas en los capítulos 9 al 14). Finalmente, indica el nivel de significación, como por ejemplo "$p<0{,}05$" ó "$p<0{,}01$".

Reber y Kotovsky (1997), en un estudio acerca de la resolución de problemas, describieron uno de sus resultados comparando un grupo específico de participantes dentro del grupo de control general. Lo hicieron de la siguiente manera: "Este grupo necesitó un promedio de 179 movimientos para resolver el rompecabezas, mientras que el resto de los participantes de control necesitaron un promedio de 74 movimientos, $t(19) = 3{,}31$, $p<0{,}01$" (p. 183). Cuando los investigadores escriben "$p<0{,}01$", quieren decir que si la hipótesis nula fuera verdadera la probabilidad de sus resultados sería menor a 0,01 (1%).

Como observamos anteriormente, la mayoría de los psicólogos sostienen que un resultado debería tener una probabilidad menor al 5% ($p<0{,}05$) para ser significativo. Por el otro lado, si un resultado se acerca, pero realmente no llega al nivel de significación del 5%, de todos modos puede ser informado como "tendencia casi significativa" o como "casi significativo", con "$p<0{,}10$". Aun cuando un resultado sea claramente no significativo, de todos modos puede figurar el nivel *p* real (por ejemplo, "$p = 0,27$"), o se puede utilizar la abreviatura NS, (por "no significativo"). Además, por lo general también se indicará si se utilizó una prueba de una cola. Como dijimos anteriormente, salvo que se indique lo contrario, al leer publicaciones científicas se supone que se utilizó una prueba de dos colas.

Aun cuando el investigador haya elegido previamente el nivel de significación, como por *ejemplo* 0,05, pueden indicarse los casos en los que los resultados cumplen con estándares más rigurosos. (Se supone que esto debe impresionar al lector). Por eso, en la misma publicación podemos encontrarnos, por ejemplo, con resultados en los cuales se indica "$p<0,05$", y en otros "$p<0,01$", e incluso en otros "$p<0,001$".

Finalmente, en muchos casos los resultados de la prueba de hipótesis se muestran sólo como asteriscos en una tabla de resultados. En esas tablas, un resultado con un asterisco es significativo, mientras que un resultado sin asterisco no lo es. Por ejemplo, la tabla 6-1 presenta los resultados de una parte de un estudio realizado por Stipek y Ryan (1997), el cual compara alumnos de jardín de infantes de condición socioeconómica baja con otros económicamente privilegiados. La tabla nos proporciona las cifras correspondientes a variables medidas a través de la observación de niños en el aula, e incluye las medias, los desvíos estándar y el estadístico F (una indicación del procedimiento utilizado en este estudio para probar la significación, procedimiento que trataremos en los capítulos 11 al 13). Lo que resulta importante observar en la tabla, para los fines que estamos tratando, son los asteriscos (y las notas correspondientes en la parte inferior de la tabla) que indican los niveles de significación de las distintas medidas. Podemos ver, por ejemplo, con respecto al deseo de demostrar los logros, que los niños de bajo nivel socioeconómico ($M = 0,20$) marcaron registros significativamente mayores a los de los niños económicamente privilegiados ($M = 0,04$). En el caso de "sonríe después de terminar la tarea" el patrón fue lo contrario.

No obstante, podemos observar que en cuanto a hacer comparaciones sociales positivas no hubo diferencias significativas entre los grupos (las *medias fueron 0,71 y 0,61, pero no fueron lo* suficientemente diferentes como para resultar significativas en este estudio). Por eso, no podemos concluir que en alumnos de jardín de infantes una mala condición económica tenga alguna relación con haber realizado comparaciones sociales positivas. También sería equivocado llegar a la conclusión de que una mala situación económica no tiene ninguna relación con realizar comparaciones sociales positivas. Como dijimos anteriormente, cuando un resultado no es lo suficientemente fuerte como para que se rechace la hipótesis nula, normalmente la mejor conclusión es que los resultados no son concluyentes.

Cabe mencionar que en todos estos ejemplos, los investigadores por lo general no hacen explícita la hipótesis de investigación o la hipótesis nula, *ni* tampoco describen ninguno de los otros pasos del proceso en detalle. Se supone que el lector comprende perfectamente todo el proceso.

RESUMEN

La idea básica de una prueba de hipótesis es analizar la probabilidad de que el resultado de un estudio pudiera haber sucedido aun si la situación real implicase que el procedimiento experimental no produjo ninguna diferencia. Si la probabilidad es baja, se rechaza el escenario de la no diferencia, y se sostiene la teoría a partir de la cual surgió el procedimiento experimental. La expectativa de una diferencia es la hipótesis de investigación, y la situación imaginaria en la que no existe ninguna diferencia se denomina hipótesis nula. Cuando un resultado fuera muy inverosímil, si la hipótesis nula fuera verdadera, entonces se rechaza la hipótesis nula y se sostiene la hipótesis de

Tabla 6-1.
Valores medios de variables observadas en clase, relacionadas con la motivación según la situación socioeconómica.

	Condición socioeconómica baja		*Privilegiados*		
Variable de motivación	*M*	*SD*	*M*	*SD*	*F*(1, 195)
Desea demostrar sus logros	0,20	0,51	0,04	0,20	9,94**
Sonríe después de terminar la tarea	0,14	0,42	0,05	0,22	4,49*
Comparación social positiva	0,71	0,45	0,64	0,48	0,01
Comparación social negativa	0,12	0,34	0,36	0,48	21,24****
Comentarios sobre competencia	4,14	1,83	5,74	1,78	25,39****
Busca ayuda	0,01	0,10	0,09	0,33	5,14*
Incumplimiento	0,12	0,35	0,13	0,53	0,07
Disciplina	0,10	0,30	0,16	0,47	2,26
Tristeza	1,03	0,17	1,02	0,14	0,15
Aburrimiento	1,05	0,21	1,29	0,46	25,29***
Frustración	1,03	0,17	1,03	0,17	0,02
Nivel de esfuerzo	1,34	0,93	1,36	0,95	0,28

$^*p < 0{,}05$; $^{**}p < 0{,}01$; $^{***}p < 0{,}001$; $^{****}p < 0{,}0001$.

Fuente: Stipek, D. J., & Ryan, R. H. (1997), tab. 4. "Alumnos de jardín de infantes con desventajas económicas: listos para aprender pero con un camino más largo para recorrer". *Psicología del Desarrollo [Developmental Psychology]*, 33, 711–723. Copyright, 1997, por la Asociación Americana de Psicología [American Psychological Association]. Reimpreso con autorización.

investigación. Si los resultados obtenidos no son muy extremos, se dice que el estudio no fue concluyente.

Los psicólogos usualmente consideran un resultado como muy extremo si presenta menos de un 5% de posibilidades, aunque algunas veces se utiliza un corte más riguroso, del 1%. Estos porcentajes pueden aplicarse a la probabilidad de que un resultado sea extremo en una dirección predicha (prueba direccional o de una cola), o a la probabilidad de que sea extremo en cualquiera de las dos direcciones posibles (prueba no direccional o de dos colas). Para aplicar una política más conservadora, los psicólogos utilizan con frecuencia las pruebas de dos colas aun cuando ya tengan una predicción específica.

El proceso de prueba de hipótesis involucra cinco pasos:

1. Replantear el problema en función de la hipótesis de investigación e hipótesis nula de las poblaciones.
2. Determinar las características de la distribución comparativa.
3. Determinar el punto muestral de corte en la distribución comparativa, a partir del cual debería rechazarse la hipótesis nula.
4. Determinar el valor muestral en la distribución comparativa.
5. Comparar los valores de los pasos 3 y 4 para decidir si se rechaza o no la hipótesis nula.

Una gran controversia ha surgido recientemente con respecto a las pruebas de significación. Los críticos han planteado cuestiones sobre la lógica básica de estas pruebas. Sin embargo, la principal crítica plantea que, con mucha frecuencia, las pruebas son mal utilizadas. Una manera que tienen los investigadores para utilizar inadecuadamente las pruebas es interpretando que el no rechazo de la hipótesis nula implica sostenerla.

En general, las publicaciones científicas informan los resultados de la prueba de hipótesis indicando si fueron o no significativas y mostrando el nivel de corte de la probabilidad (generalmente del 5% ó 1%) según el cual fue tomada la decisión.

Términos Clave

- Distribución comparativa.
- Niveles convencionales de significación ($p<0{,}05$, $p<0{,}01$).
- Punto muestral de corte .
- Hipótesis direccional.
- Prueba de hipótesis.
- Hipótesis no direccional.
- hipótesis nula.
- Prueba de una cola.
- Hipótesis de investigación.
- Estadísticamente significativo.
- Prueba de dos colas.

Ejercicios

Los ejercicios implican la realización de cálculos (con la ayuda de una calculadora). La mayoría de los problemas estadísticos reales se resuelven por computadora, pero aunque exista la posibilidad de utilizarla, es conveniente realizar estos ejercicios manualmente para incorporar el método de trabajo.

Para adquirir práctica en la utilización de una computadora, para resolver problemas estadísticos, se puede utilizar la sección de computación de cada capítulo, publicada en la *Guía de estudio y libro de tareas de computación para el alumno [Student's Study Guide and Computer Workbook]* que acompaña este libro.

Todos los datos de esta sección son ficticios (a menos que se especifique lo contrario).

Las respuestas a los ejercicios de la serie I se encuentran al final del libro.

SERIE I

1. Defina los siguientes términos utilizando sus propias palabras: a) hipótesis de investigación, b) hipótesis nula, c) procedimiento de prueba de hipótesis, d) distribución comparativa, e) nivel de significación 0,05, y f) prueba de una cola.

2. Lea atentamente los tres puntos que aparecen a continuación y, a) indique cuáles son las dos poblaciones que se comparan, b) establezca la hipótesis de investigación, c) establezca la hipótesis nula y d) determine si se debería utilizar una prueba de una o dos colas y por qué.

I) Los niños canadienses hijos de bibliotecarios ¿tienen una mayor habilidad para la lectura que los niños canadienses en general?

II) El nivel de ingreso de los residentes de determinada ciudad ¿es diferente del nivel de ingresos de los habitantes de la región?

III) Las personas que han sufrido la experiencia de un terremoto ¿tienen más o menos confianza en sí mismas que la población en general?

3. Basándose en la información obtenida de cada uno de los siguientes estudios, determine si se rechaza o no la hipótesis nula. En cada caso, determine: a) la puntuación Z de corte en la distribución comparativa, a partir de la que debería rechazarse la hipótesis nula; b) la puntuación Z muestral en la distribución comparativa, y c) su conclusión. (Suponga que todas las poblaciones están normalmente distribuidas).

Estudio	Población μ	σ	Registro muestral	p	Colas de la prueba
A	10	2	14	0,05	1 (predicción alta)
B	10	2	14	0,05	2
C	10	2	14	0,01	1 (predicción alta)
D	10	2	14	0,01	2
E	10	4	14	0,05	1 (predicción alta)
F	10	1	14	0,01	2
G	10	2	16	0,01	2
H	12	2	16	0,01	2
I	12	2	8	0,05	1 (predicción baja)

4. Una psicóloga interesada en los sentidos del gusto y del olfato ha realizado una serie extensiva de estudios en los que hace probar a alumnos universitarios 20 tipos de alimentos diferentes (damasco, chocolate, cereza, café, ajo, y otros). Cada alimento se suministra en forma de gota que se vierte sobre la lengua. De toda la población de alumnos de la universidad, la cantidad media que los alumnos pueden identificar correctamente entre estos 20 alimentos es 14, con un desvío estándar de 4. (Supongamos que todos los alumnos de esa facultad son examinados como parte de una investigación médica al comienzo de cada año). La psicóloga tiene razones para creer que la precisión de las personas, en esta prueba, está más relacionada con el olfato que con el gusto. Por lo tanto, establece procedimientos especiales que impiden utilizar el sentido del olfato durante la prueba. Luego, la psicóloga prueba el procedimiento en un alumno seleccionado al azar. El alumno identifica correctamente sólo 5 alimentos. Utilizando el nivel de significación 0,05, ¿qué conclusión debería sacar la investigadora? Resuelva este problema utilizando explícitamente los cinco pasos de la prueba de hipótesis. Luego explique su respuesta a alguien que nunca ha asistido a un curso de estadística (pero que está familiarizado con los conceptos de media, desvío estándar y puntuaciones Z).

5. Un psicólogo está trabajando con personas que han tenido un tipo particular de cirugía mayor. El psicólogo propone la teoría de que una persona se recuperará más rápido de la operación si los amigos y la familia están en la habitación con el paciente durante las primeras 48 horas siguientes a la operación. Se sabe (en este ejemplo ficticio) que el tiempo de recuperación está distribuido normalmente con una media de 12 días y un desvío estándar de 5 días. El procedimiento se prueba con un paciente seleccionado al azar, que se recupera en 18 días. Utilizando el nivel de significación 0,01, ¿qué conclusión debería sacar el investigador? Resuelva este problema utilizando explícitamente los cinco pasos de la prueba de hipótesis. Luego explique su respuesta a alguien que nunca ha asistido a un curso de estadística (pero que está familiarizado con los conceptos de media, desvío estándar y puntuaciones Z).

6. Robins y John (1997) realizaron un estudio sobre el narcisismo (egolatría), en el que se comparaban individuos que habían tenido valores altos con individuos que habían obtenido valores bajos (con ítems tales como: "Si yo gobernara el mundo, éste sería un lugar mejor"). También realizaban algunas otras preguntas, incluyendo un ítem en el que se preguntaba a los participantes cuántas veces se miraban al espejo en un día típico. Al informar sobre los resultados, los investigadores observaron:

> "... tal como se había predicho, los individuos con un alto grado de narcisismo informaron que se miraban al espejo con más frecuencia que los individuos con un bajo nivel de narcisismo (*M*s = 5,7 vs 4,8)... p<0,05" (p. 39).

Explique este resultado a una persona que nunca ha asistido a un curso de estadística. (Concéntrese en el significado del resultado en cuanto a la lógica general de la prueba de hipótesis y a la significación estadística).

SERIE II

1. Enumere los pasos del proceso de prueba de hipótesis y explique el procedimiento y los fundamentos de cada uno.

2. Para cada uno de los puntos que se detallan a continuación, a) indique cuáles son las dos poblaciones que se comparan, b) determine la hipótesis de investigación, c) determine la *hipótesis nula* y d) explique si se debería utilizar una *prueba de una o dos colas* y por qué.

I) En un experimento, se dan instrucciones a los participantes para que resuelvan un problema concentrándose en los detalles. ¿Es diferente la velocidad con la que resuelven el problema las personas que han recibido tales instrucciones, en comparación con las personas a las que no se les ha dado ninguna instrucción especial?

II) A partir de informes antropológicos en los que se registra la condición social de la mujer en una escala de 10 puntos, se conocen la media y el desvío estándar en muchas culturas. Se descubre una nueva cultura en la que existe una organización familiar inusual. También se clasifica la condición social de la mujer en esta cultura. ¿Las culturas con una or- organización familiar inusual brindan a la mujer una condición social más elevada que las culturas en general?

III) ¿Las personas que viven en grandes ciudades sufren más enfermedades relacionadas con el estrés que las personas en general?

3. A partir de la información correspondiente a cada uno de los siguientes estudios, determine si se rechaza o no la hipótesis nula. En cada caso, establezca a) la puntuación Z de corte en la distribución comparativa a partir de la cual debería rechazarse la hipótesis nula; b) la puntuación Z muestral en la distribución comparativa, y c) su conclusión. (Suponga que todas las poblaciones están normalmente distribuidas).

Estudio	Población		Observación muestral	p	Colas de la prueba
	μ	σ			
A	100,0	10,0	80	0,05	1 (predicción baja)
B	74,3	11,8	80	0,01	2
C	16,9	1,2	80	0,05	1 (predicción baja)
D	88,1	12,7	80	0,05	2

4. Un investigador ha descubierto que ciertos sonidos hacen a las ratas mucho más agresivas, y predice que los sonidos también disminuirán sus desempeños en cuanto a tareas de aprendizaje. Supongamos que se sabe que una rata promedio, ordinaria, puede aprender a correr correctamente en un determinado laberinto en 18 pruebas, con un desvío estándar de 6. El investigador, entonces, prueba una rata ordinaria en el laberinto, pero haciéndole escuchar el sonido. La rata necesita 38 intentos para aprender el laberinto. Utilizando el nivel 0,05, ¿qué conclusión debería sacar el investigador? Resuelva este problema utilizando explícitamente los cinco pasos de la prueba de hipótesis. Luego explique su respuesta a alguien que nunca ha asistido a un curso de estadística (pero que está familiarizado con los conceptos de media, desvío estándar y puntuación Z).

5. Un psicólogo especializado en temas de familia ha desarrollado un elaborado programa de capacitación para contribuir a la adaptación de hombres sin hijos casados con mujeres con hijos adolescentes. Supongamos que se sabe, a partir de investigaciones previas, que estos hombres, un mes después de mudarse con la nueva esposa y sus hijos, sufren un nivel de estrés de 85 con un desvío estándar de 15. Como experimento piloto, se prueba el programa de capacitación en un hombre seleccionado al azar de entre todos aquellos en determinada ciudad que, durante el mes anterior, se habían casado con una mujer con un hijo adolescente. Después del programa de capacitación, el nivel de estrés de ese hombre es 60. Utilizando el nivel 0,05, ¿qué conclusión debería sacar el investigador? Resuelva este problema utilizando explícitamente los cinco pasos de la prueba de hipótesis. Luego explique su respuesta a alguien que nunca ha asistido a un curso de estadística (pero que está familiarizado con los conceptos de media, desvío estándar y puntuación Z).

6. En una publicación acerca de las campañas en contra del tabaco, realizado en Massachussetts en 1993 y 1995, Siegel y Biener

(1997) exponen los resultados de una encuesta sobre el consumo de tabaco y las distintas actitudes. La tabla 6-2 muestra los resultados de esta encuesta. Concentrándose sólo en la primera línea (porcentaje que fuma > 25 por día), explique qué significa el resultado a una persona que nunca ha asistido a un curso de estadística. (Concéntrese en el significado del resultado en cuanto a la lógica general de la prueba de hipótesis y a la significación estadística).

Tabla 6-2.
Algunos indicadores del cambio en el consumo de tabaco, exposición al ETS3, y actitudes del público hacia las políticas de control del tabaco, Massachussetts, 1993-1995.

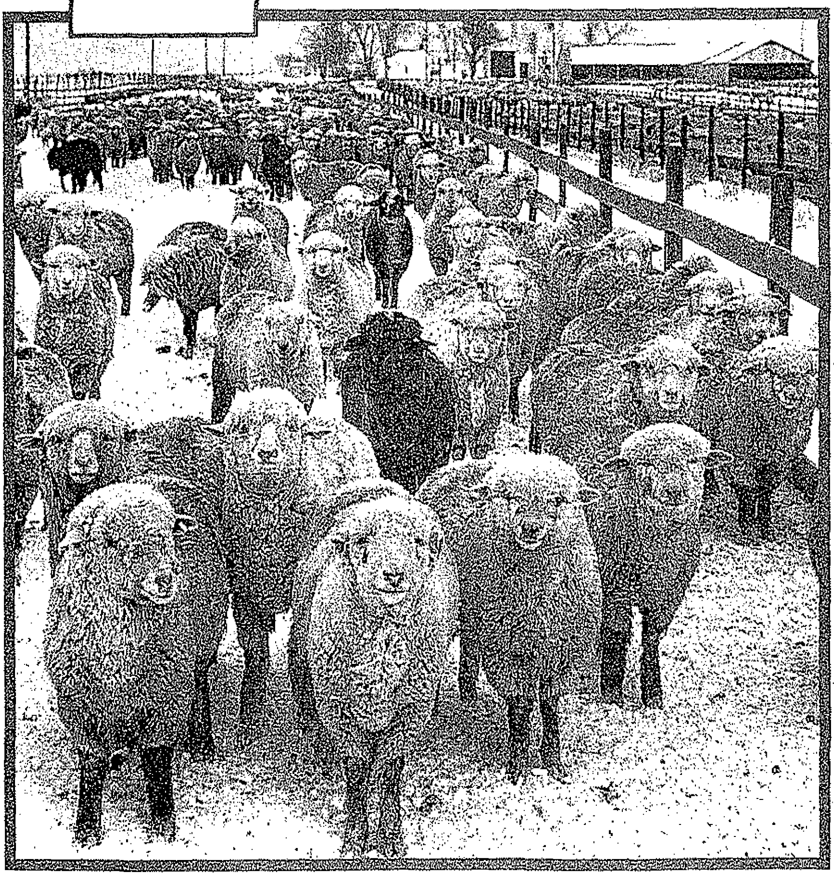
	1993	1995
Comportamiento de fumadores adultos		
Porcentaje que fuma > 25 cigarrillos diarios	24	10*
Porcentaje que fuma < 15 cigarrillos diarios	31	49*
Porcentaje que fuma antes de transcurridos 30 minutos de despertarse	54	41
Exposición al humo de tabaco en el ambiente		
Porcentaje de trabajadores que informan sobre un lugar de trabajo en el que no se fuma	53	65*
Media de horas de exposición al ETS en el trabajo durante la semana anterior	4,2	2,3*
Porcentaje de hogares en los que está prohibido fumar	41	51*
Actitudes hacia las políticas de control del tabaco		
Porcentaje que apoya un mayor aumento de impuestos al tabaco asignando los fondos al control del tabaco	78	81
Porcentaje que cree que la exposición al ETS es perjudicial	90	84
Porcentaje que apoya la prohibición de las máquinas expendedoras	54	64*
Porcentaje que apoya la prohibición del patrocinio de deportes y eventos culturales por parte de las compañías de tabaco	59	53*

Fuente: Biener y Roman, 1996.

*$p < 0,05$.

Fuente: Siegel, M., & Biener, L. (1997), tab. 4. "Evaluación del impacto de las campañas estatales contra el tabaco: programas de control del tabaco de Massachusetts y California". *Revista Científica de Asuntos Sociales [Journal of Social Issues]*, 53, 147-168.

7 Pruebas de hipnósis con medias muestrales



Descripción del capítulo

- La distribución de medias.
- Creación de una distribución de medias.
- Características de una distribución de medias.
- Prueba de hipótesis sobre la distribución de medias.
- Estimación e intervalos de confianza.
- Controversias y limitaciones: ¿Intervalos de confianza o pruebas de significación?
- Desvío estándar de la distribución de la media muestral, pruebas de hipótesis sobre la media e intervalos de confianza según se describen en publicaciones científicas.
- Resumen.
- Términos clave.
- Ejercicios.

En el capítulo 6 presentamos la lógica básica de la prueba de hipótesis. Utilizamos como ejemplos estudios en los que la muestra estaba formada por un sólo individuo. Sin embargo, como señalamos anteriormente, en la práctica, la investigación psicológica usualmente utiliza muestras integradas por muchos individuos. En este capítulo nos basamos en lo aprendido hasta ahora y analizamos la prueba de hipótesis con muestras de más de un individuo, lo cual requiere, principalmente, analizar con cierto detalle lo que denominamos distribución de medias.

LA DISTRIBUCIÓN DE MEDIAS

La prueba de hipótesis en condiciones normales de investigación, cuando se analiza una muestra formada por muchos individuos, es exactamente igual a lo que hemos aprendido en el capítulo 6, con una importante excepción. Cuando hay más de una persona en la muestra surge un problema específico en el paso 2, al determinar las características de la distribución comparativa. El problema es que el valor muestral que nos interesa es la media del grupo de valores. Las distribuciones comparativas que hemos estado analizando hasta ahora han sido distribuciones poblacionales de valores individuales (por ejemplo, las edades en que cada bebé en particular comienza a caminar o la población de valores individuales a partir de un cuestionario para medir el nivel de felicidad). Comparar la media de una muestra de, digamos, 50 individuos con una distribución de valores individuales constituye una comparación desigual, como comparar manzanas y naranjas. En cambio, cuando lo que nos interesa es la media de una muestra de 50, necesitamos una distribución comparativa formada por medias de muestras de 50 valores. A esta distribución comparativa la denominaremos **distribución de medias.**

Para expresarlo más formalmente, una distribución de medias es una distribución formada por las medias de cada una de las numerosas muestras del mismo tamaño seleccionadas al azar entre la misma población de individuos. (Los estadísticos también llaman a esta distribución de medias una "**distribución en el muestreo de la media**"; sin embargo, en este libro utilizamos el término **distribución de medias** para que quede claro que estamos hablando de poblaciones y no de muestras o distribuciones de frecuencias de una muestra).

La distribución de medias es la distribución comparativa adecuada cuando la muestra está formada por más de una persona. Por eso, en la mayoría de las investigaciones resulta necesario determinar las características de esa distribución para poder realizar el paso 2 del procedimiento de prueba de hipótesis.

CREACIÓN DE UNA DISTRIBUCIÓN DE MEDIAS

Podremos comprender mejor la idea de una distribución de medias analizando cómo construir tal distribución a partir de una distribución ordinaria de individuos. Supongamos que nuestra población estaba compuesta por alumnos de grados elementales e inferiores –de determinada región– cuya población total es de 90.000 niños. Supongamos, además (para que el ejemplo sea simple), que hay exactamente 10.000 niños en cada grado, desde el primero hasta el noveno. La distribución de población sería rectangular, con una media de 5, una varianza de 6,67 y un desvío estándar de 2,58 (*véase* figura 7-1).

Supongamos que luego escribiéramos el grado al que pertenece cada niño en una pelotita de ping pong y que pusiéramos las 90.000 pelotitas plásticas en un recipiente gigante. El recipiente contendría 10.000 pelotitas con un número 1 escrito en ellas, 10.000 con un número 2, y así sucesivamente. Mezclamos las pelotitas en el recipiente, y luego extraemos dos, es decir, se extrae una muestra aleatoria de dos pelotitas. Supongamos que una pelotita tiene un número 2 y la otra tiene un número 9. En ese caso, el grado medio de la muestra formada por el grado al que pertenecen dos de los niños es 5,5, es decir, el promedio de 2 y 9. Entonces, volvemos a poner las pelotitas en el recipiente, mezclamos todas las pelotitas, y seleccionamos otra vez dos pelotitas. Puede ser que esta vez extraigamos dos cuatros, siendo 4 la media de la segunda muestra. Después volvemos a realizar el procedimiento, y en esa oportunidad extraemos un 2 y un 7, siendo la media 4,5. Hasta aquí tenemos tres medias: 5,5, 4 y 4,5.

Los tres números mencionados en el párrafo anterior (cada uno de ellos es la media de una muestra formada por los grados a los que pertenecen dos niños de escuela) pueden considerarse una pequeña distribución en sí misma. La media de esta pequeña distribución de tres números es 4,67 (la suma de 5,5, 4 y 4,5, dividida por 3); la varianza de esa distribución es 0,39 (la varianza de 5,5, 4 y 4,5) y el desvío estándar es 0,62 (la raíz cuadrada de 0,39). La figura 7-2 representa un histograma de esta distribución de tres medias.

Si continuáramos con el proceso, el histograma de medias continuaría creciendo. La figura 7-3a representa un ejemplo después de seleccionar 10 muestras aleatorias de dos pelotitas cada

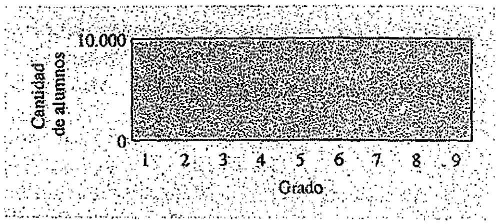


Figura 7-1. Distribución del grado de 90.000 escolares (datos ficticios).

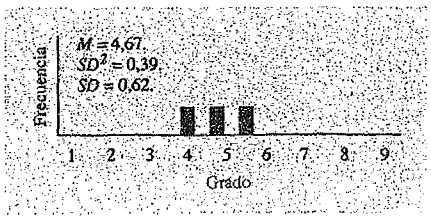


Figura 7-2. Distribución de medias de tres muestras aleatorias de los grados a los que pertenecen dos escolares, extraídas de una población conformada por los grados a los que concurren 90.000 escolares (datos ficticios).

una. La figura 7-3b representa el histograma de la distribución de medias después de seleccionar 20 muestras aleatorias de dos registros cada una. Después de seleccionar 100 muestras aleatorias, el histograma de la distribución de medias podría verse como la figura 7-3c; después de 1.000, como la figura 7-3d. (En realidad, en lugar de utilizar 90.000 pelotitas de ping pong y un recipiente gigante, creamos los histogramas de la figura 7-3 por medio de una computadora que realizó las selecciones aleatorias).

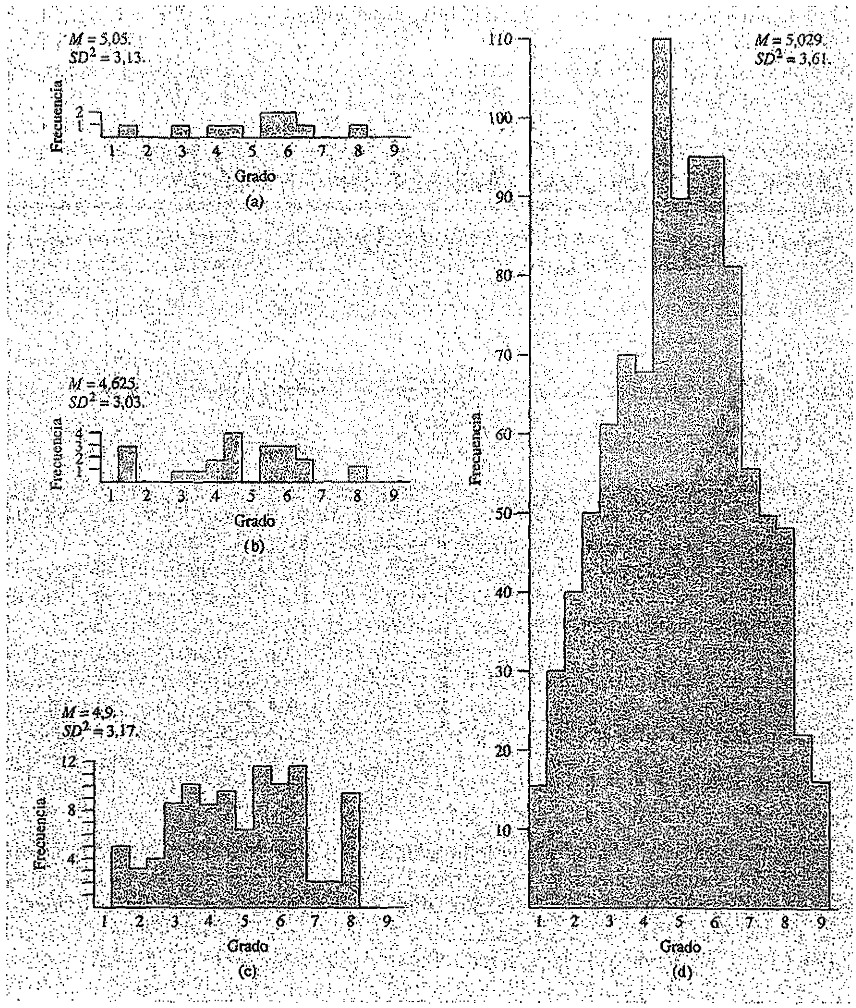


Figura 7-3. Distribuciones de medias de muestras aleatorias de dos pelotitas cada una, extraídas de una población de 90.000 pelotitas, de las cuales, cada 10.000, llevaban uno de los números del 1 al 9. Las cantidades de medias muestrales que incluye cada distribución son (a) 10 medias muestrales, (b) 20 medias muestrales, (c) 100 medias muestrales y (d) 1.000 medias muestrales. (El muestreo real fue simulado por computadora).

En la práctica, los investigadores casi nunca tienen la oportunidad de seleccionar muchas muestras diferentes de una población. Lleva mucho trabajo poder lograr una sola muestra y estudiar a la gente que la conforma. Sin embargo, afortunadamente podemos determinar las características de una distribución de muestras en forma directa utilizando algunas reglas simples, sin necesidad de seleccionar siquiera una sola muestra. La única información que necesitamos es: a) características de la distribución de la población de individuos y b) tamaño de cada muestra. (Por ahora no nos preocuparemos por cómo podríamos conocer las características de la población de individuos). El trabajoso método de construir una distribución de medias en la forma que acabamos de hacerlo y el método conciso que aprenderemos muy pronto, tienen el mismo resultado. Hemos analizado el proceso de ese modo meticuloso sólo porque ésto ayuda a comprender el concepto de una distribución de medias.

CARACTERÍSTICAS DE UNA DISTRIBUCIÓN DE MEDIAS

Observemos tres temas relacionados con la distribución de medias que construimos según nuestro ejemplo (según aparece en la figura 7-3):

1. La media de la distribución de medias resultó ser aproximadamente igual a la media de la población original formada por los grados individuales, de la cual se extrajeron las muestras (en ambos casos la media fue 5).
2. La dispersión de la distribución de medias resultó ser menor que la dispersión de la distribución poblacional de la cual se extrajeron las muestras.
3. La forma de la distribución de medias resultó ser aproximadamente normal (o al menos unimodal y simétrica).

Sucede que los dos primeros de los tres puntos anteriores son ciertos para todas las distribuciones de medias, y el tercero es cierto para la mayoría de ellas.

Estos tres vínculos entre la distribución de medias y la población de individuos constituyen los fundamentos de una serie de reglas simples que nos permiten determinar la media, la varianza y la forma de una distribución de medias sin tener que escribir en pelotitas plásticas ni seleccionar interminables muestras. Las tres reglas, que pronto volveremos a analizar, se basan en el **teorema del límite central**, un principio fundamental en estadística matemática que ya hemos mencionado en el capítulo 5.

Analicemos ahora las tres reglas.

Regla 1: determinación de la media de una distribución de medias

La primera regla establece que **la media de una distribución de medias** es igual a la media de la población de individuos de la cual se extrajeron las muestras. Cada muestra se basa en valores seleccionados al azar de la población de individuos. Así, a veces, la media de una muestra será mayor que la media de toda la población de individuos y, otras veces la media de una muestra será menor que la media de toda la población de individuos. Sin embargo, no existe razón para que las medias de estas muestras tiendan a ser consistentemente mayores o menores, en su conjunto, que la media de la población de individuos. Si se seleccionan suficientes muestras, las medias altas y las medias bajas se equilibran entre sí.

Podemos ver en la figura 7-3 que con un gran número de muestras, la media de la distribución de medias se torna muy similar a la media de la población de individuos, que en este caso era 5. Si hubiéramos mostrado un ejemplo con 10.000 medias muestrales, hubiéramos estado aún más cerca del 5. Se puede probar matemáticamente que si tomáramos una cantidad infinita de muestras, la media de la distribución de medias de estas muestras resultaría ser exactamente igual a la media de la distribución de individuos.

Regla 2: determinación de la varianza de una distribución de medias

La figura 7-3 también muestra que una distribución de medias estará menos dispersa que la población de individuos de la que se extrajeron las muestras. La razón es la siguiente: en una muestra aleatoria cualquier valor tiene posibilidades de ser seleccionado, incluso un valor extremo, pero la posibilidad de seleccionar dos valores extremos en la misma muestra aleatoria es menor. Más aún, para crear una media muestral extrema, los dos valores extremos tendrían que serlo en la misma dirección. Por lo tanto, aumentar las cantidades produce un efecto moderador. En cualquier muestra, los extremos tienden a ser equilibrados por los valores centrales o por extremos en la dirección opuesta. Esto hace que cada media muestral tienda hacia los valores centrales y se aleje de los extremos. Habiendo menos medias extremas, la varianza de las medias es menor.

Analicemos nuestro ejemplo. En la población hay muchos unos y nueves que crean una cantidad considerable de dispersión. Es decir, si extrajéramos muestras formadas por un sólo valor, aproximadamente una novena parte de las veces obtendríamos un 1, y aproximadamente una novena parte de las veces obtendríamos un 9. Sin embargo, si seleccionáramos muestras de dos registros por vez, obtendríamos una muestra con una media de 1 (es decir, en la que ambas pelotitas fueran unos) o una media de 9 (en la que **ambas** pelotitas fueran nueves) con mucha menos frecuencia, ya que hay más chances de obtener dos pelotitas que promedien un valor medio como el 5. (Esto se debe a que existen varias combinaciones que podrían dar ese resultado: un 1 y un 9, un 2 y un 8, un 3 y un 7, un 4 y un 6 y dos 5).

Cuantos más valores haya en cada muestra, menos dispersa será la distribución de medias de esas muestras, ya que, con varios valores en cada muestra, es aún más extraño que los valores extremos de cualquier muestra no sean equilibrados por valores centrales o extremos en otra dirección. Con respecto al ejemplo de las pelotitas de plástico, vimos que era bastante improbable obtener una media de 1 tomando muestras de dos pelotitas a la vez. Si seleccionáramos tres pelotitas a la vez, obtener una muestra con una media de 1 (las tres pelotitas deberían ser unos) sería aún menos probable. Se hace cada vez más probable la obtención de medias con valores centrales.

En nuestro ejemplo, utilizando muestras de dos pelotitas cada una, la varianza de la distribución de medias será aproximadamente 3,33, que equivale a la mitad de la varianza de la población de pelotitas individuales, que era de 6,67. Si hubiéramos creado una distribución de medias utilizando muestras de tres pelotitas cada una, la varianza de la distribución de medias hubiera sido de 2,22, es decir, un tercio de la varianza de la población de individuos. Si hubiéramos seleccionado al azar cinco pelotitas para cada muestra, la varianza de la distribución de medias hubiera sido un quinto de la varianza de la población de individuos.

Los ejemplos anteriores cumplen una regla general, la segunda regla con respecto a la distribución de medias: **la varianza de una distribución de medias** es la varianza de la distribución de la población de individuos dividida por el tamaño de cada una de las muestras seleccionadas. Esta regla se mantiene en todas las situaciones y se puede probar matemáticamente.

A continuación mostramos la fórmula que representa la regla para calcular la varianza de la distribución de medias:

$$\sigma_M^2 = \frac{\sigma^2}{N} \tag{7-1}$$

En la fórmula anterior, σ_M^2 es la varianza de la distribución de medias, σ^2 es la varianza de la población de individuos, y N es el tamaño de cada muestra.

En nuestro ejemplo, la varianza de la población de grados individuales era 6,67, y había dos niños de escuela por muestra. La varianza de la distribución de medias se calcula del siguiente modo:

$$\sigma_M^2 = \frac{\sigma^2}{N} = \frac{6{,}67}{2} = 3{,}34$$

Para utilizar un ejemplo diferente, supongamos que una población de individuos tuviera una varianza de 400, y quisiéramos saber la varianza de una distribución de medias de 25 valores observados:

$$\sigma_M^2 = \frac{\sigma^2}{N} = \frac{400}{25} = 16$$

El **desvío estándar de una distribución de medias** es la raíz cuadrada de la varianza de la distribución de medias. Se representa bajo la fórmula:

$$\sigma_M = \sqrt{\sigma_M^2} = \sqrt{\frac{\sigma^2}{N}} \qquad (7\text{-}2)$$

En la fórmula anterior, σ_M es el desvío estándar de la distribución de medias.

Algunas veces, esta fórmula se manipula algebraicamente para destacar la relación entre el desvío estándar de la población de individuos y el desvío estándar de la distribución de medias:

$$\sigma_M = \frac{\sigma}{\sqrt{N}} \qquad (7\text{-}3)$$

Debido a su importancia en la prueba de hipótesis, a veces se denomina al desvío estándar de la distribución de medias con un nombre especial propio: **error estándar de la media** o, para abreviar, **error estándar**. Este nombre representa la medida de "error" típica de determinadas medias muestrales como estimaciones de la media de la población de individuos. Es decir, el error estándar de la media nos indica cuánto se desvían de la media poblacional las medias particulares de la distribución de medias. Al final del capítulo seguiremos tratando este tema en nuestra exposición sobre intervalos de confianza.

Regla 3: forma de una distribución de medias

No importa cuál sea la forma de la distribución original de valores individuales, ya que la distribución de medias siempre tiende a ser unimodal y simétrica. En el ejemplo de los grados escolares, la distribución poblacional de los grados de los alumnos era rectangular (cada valor tenía la misma frecuencia). Sin embargo, la **forma de la distribución de medias** era semejante a la de una campana unimodal y simétrica. Si en nuestro ejemplo de la figura 7-3 hubiéramos seleccionado muchas más de 1.000 muestras, la forma habría sido mucho más claramente unimodal y simétrica.

Una distribución de medias tiende a ser unimodal debido al mismo proceso básico de los extremos equiparándose entre sí que observamos al tratar el tema de la varianza: cuando se trata de medias es más posible que se den los valores centrales, y menos posible que se den las medias extremas. La distribución tiende a ser simétrica porque la falta de simetría (asimetría) es causada principalmente por extremos, y al haber menos extremos hay menos asimetría. En nuestro ejemplo de los grados escolares, la distribución de medias que creamos resultó tan claramente simétrica debido a que la distribución poblacional de los grados individuales era simétrica. Si la distribución de valores de la población de individuos hubiera sido asimétrica hacia un lado, la distribución de medias hubiera sido asimétrica también, pero no tanto.

Cuantos más valores haya en cada muestra, más semejante será la distribución de medias a la distribución normal. Por lo tanto, la tercera regla establece que con muestras de 30 ó más valores, aun con una población de individuos no normal, la distribución de medias se aproximará mucho a una distribución normal y los porcentajes en la tabla de áreas bajo la curva normal serán extremadamente precisos.[1,2] Además, siempre que la distribución de valores de la población de individuos sea normal, la distribución de medias será normal, sin importar la cantidad de valores que incluya cada muestra.

Resumen de las reglas para la determinación de las características de una distribución de medias

A continuación resumimos las tres reglas:

1. La media de una distribución de medias es igual a la media de la distribución poblacional de observaciones individuales.
2. La varianza de una distribución de medias es la varianza de la distribución de la población de observaciones individuales dividida por el tamaño de la muestra (cantidad de valores observados de cada muestra ($\sigma_M^2 = \sigma^2/N$). Su desvío estándar es la raíz cuadrada de su varianza ($\sigma_M = \sqrt{\sigma_M^2}$).
3. La forma de una distribución de medias es, al menos, aproximadamente normal si a) cada muestra incluye 30 valores o más, o bien, b) la distribución de observaciones de la población de individuos es normal. En otras circunstancias, aun presentará una tendencia a ser unimodal y aproximadamente simétrica.

Estos principios están representados gráficamente en la figura 7-4.

[1]Hemos ignorado el hecho de que una curva normal es una distribución teórica ininterrumpida. En la mayoría de los ejemplos de la vida real, los registros se ubican en intervalos específicos. Por lo tanto, una diferencia entre una curva normal y la distribución de medias de pelotitas de ping pong de nuestro ejemplo es que la curva normal es ininterrumpida. *Sin embargo, en la investigación psicológica, usualmente suponemos que, aun cuando nuestras mediciones se* realicen a través de intervalos específicos, el objeto implícito que estamos midiendo es continuo.

[2]Ya hemos analizado en el capítulo 5 el principio que establece la tendencia de la distribución de medias hacia una curva normal. Aunque aún no habíamos estudiado la distribución de medias, aun así utilizamos ese principio para explicar por qué la distribución de tantos elementos en la naturaleza siguen una curva normal. En ese capítulo lo explicamos como consecuencia de las distintas influencias que se equiparan unas a otras para hacer surgir una influencia promedio con la mayoría de los registros cerca del centro y, unos pocos, a cada extremo. Ahora hemos explicado el mismo tema utilizando la terminología de una distribución de medias. Pensemos en cualquier distribución de registros individuales en la naturaleza como representativa de una situación en la que cada registro es efectivamente un promedio de una serie aleatoria de influencias que actúan sobre ese registro individual. Analicemos la distribución del peso del canto rodado. El peso de cada piedra representa una especie de promedio de todas las diferentes fuerzas que actuaron para que ese canto rodado tenga un peso determinado.

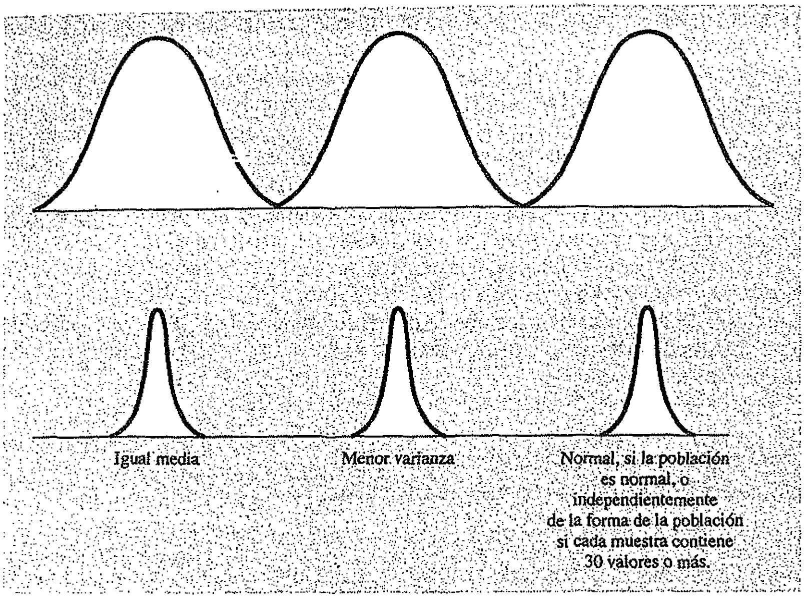


Figura 7-4. Ilustración de los principios de la relación entre la distribución de medias (curvas en la parte inferior) y la distribución de la población de observaciones individuales (curvas en la parte superior).

Ejemplo de determinación de las características de una distribución de medias

Analicemos la distribución de valores de una población de alumnos que han rendido el GRE. Supongamos que la distribución es aproximadamente normal con una media de 500 y un desvío estándar de 100. ¿Cuáles serán las características de una distribución de medias realizada con muestras de 50 alumnos cada una, seleccionados de esa población?

1. Dado que la media de la población es 500, la media de la distribución de medias también será 500.

2. La varianza de la distribución de medias es la varianza de la población de observaciones individuales dividida por la cantidad de individuos en cada muestra. Dado que el desvío estándar de la población de observaciones individuales es 100, la varianza de esa población es 10.000. La varianza de la distribución de medias es 10.000 dividido 50, es decir 200. Lo anterior se expresa bajo la siguiente fórmula,

$$\sigma_M^2 = \frac{\sigma^2}{N} = \frac{10.000}{50} = 200$$

El desvío estándar de la distribución de medias es la raíz cuadrada de la varianza de la distribución de medias: $\sqrt{200} = 14{,}14$.

3. La forma de la distribución de medias será normal. Se cumplen nuestros dos requerimientos: la distribución de valores de la población de individuos es normal y la cantidad de individuos en cada muestra es igual a 30 ó mayor. (Habría sido suficiente si se hubiera cumplido sólo uno de los requerimientos).

Otro ejemplo de determinación de las características de una distribución de medias

La Lista de Control de Adjetivos [*Adjective Check List*] (Gough & Heilbrun, 1983) es una prueba de personalidad ampliamente utilizada. La prueba está formada por una lista de adjetivos tales como **capaz, activo, atlético**, y así sucesivamente, y aquellos que realizan la prueba controlan la lista para determinar si cada adjetivo puede aplicarse a sí mismo. Una de las sub-pruebas de la Lista de Control de Adjetivos se focaliza en la agresión (adjetivos tales como **agresivo, peleador, dogmático**). La prueba ha sido aplicada a gran cantidad de personas en el pasado, y se sabe que los valores en la escala de agresión presentan una distribución asimétrica con una media de 51 y una varianza de 93 (redondeando). ¿Cuáles serán las características de una distribución de medias muestrales de esta población de individuos si cada muestra contiene 10 individuos?

1. La media de la distribución de medias será 51, la misma que la media poblacional.

2. La varianza de la distribución de medias será 93, la varianza poblacional, dividida por 10 (tamaño de cada muestra). El resultado es 9,3. Se representa bajo la fórmula:

$$\sigma_M^2 = \frac{\sigma^2}{N} = \frac{93}{10} = 9,3$$

El desvío estándar de la distribución de medias es la raíz cuadrada de 9,3, ó lo que es lo mismo, 3,05.

3. La distribución de medias no será normal porque la distribución de la población de individuos no es normal, y la cantidad de individuos por muestra es sólo 10. Sin embargo, como toda, distribución de medias, tendrá tendencia a ser unimodal y más simétrica que la distribución de la población de valores individuales.

Revisión de tres clases de distribuciones

Hemos estudiado tres diferentes tipos de distribuciones: a) la distribución de valores de una población de individuos, b) la distribución de observaciones de una determinada muestra tomada de esa población y c) la distribución de medias. La figura 7-5 ilustra estas tres distribuciones, y la tabla 7-1 las compara.

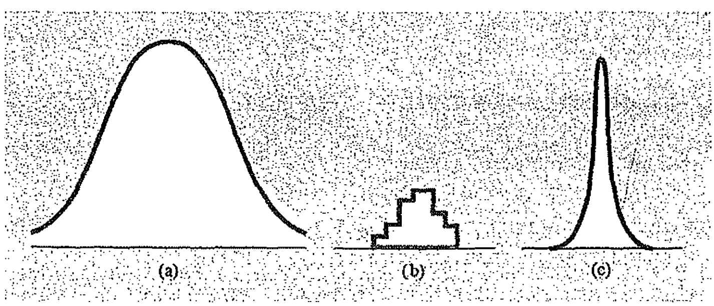


Figura 7-5. Tres tipos de distribuciones: (a) distribución de valores de una población de individuos, (b) distribución de observaciones de una determinada muestra tomada de esa población y (c) distribución de medias de todas las muestras posibles de un determinado tamaño, tomadas de esa distribución.

Tabla 7-1.
Comparación de tres tipos de distribuciones.

	Distribución poblacional	**Distribución de una muestra determinada**	**Distribución de medias**
Contenido	Valores de todos los individuos de la población.	Valores de los individuos de una sola muestra.	Medias de muestras tomadas al azar de la población.
Forma	Podría ser cualquier forma, a menudo normal.	Podría ser cualquier forma.	Normal, si la población es normal. Aproximadamente normal, si las muestras contienen ≥ 30 observaciones cada una.
Media	μ	$M = \Sigma X/N$, Calculado de las observaciones tomadas de la muestra	$\mu_M = \mu$
Varianza	σ^2	$SD^2 = \Sigma(X - M)^2/N$, Calculado de las observaciones tomadas de la muestra	$\sigma^2_M = \sigma^2/N$
Desvío estándar	σ	$SD = \sqrt{SD^2}$	$\sigma_M = \sqrt{\sigma^2_M} = \sigma/\sqrt{N}$

PRUEBA DE HIPÓTESIS CON UNA DISTRIBUCIÓN DE MEDIAS

Ahora estamos listos para analizar las pruebas de hipótesis cuando existe más de un individuo en la muestra del estudio.

La distribución de medias como distribución comparativa en la prueba de hipótesis

En esta nueva situación, la distribución de medias proporciona la conexión decisiva entre la muestra y la hipótesis nula. Supongamos que estamos estudiando una muestra de más de una persona (situación usual en las investigaciones). En ese caso, la distribución de medias es la **distribución comparativa**, la distribución cuyas características se determinan en el paso 2 del proceso de prueba de hipótesis. La distribución de medias es la distribución con la que se puede comparar la media muestral para determinar cuán verosímil es que dicha media muestral hubiera sido seleccionada si la hipótesis nula fuera verdadera.

Determinación de la puntuación Z de una media muestral en la distribución de medias

Cuando se realiza una prueba de hipótesis con una muestra de más de un individuo puede surgir cierta confusión al determinar la ubicación de la muestra en la distribución comparativa. En ese caso, lo que estamos determinando es la puntuación Z de la media muestral en la distribución de medias. (Antes determinábamos la puntuación Z de un sólo individuo en una distribución de valores de una población de individuos). El método para convertir la media muestral en una puntuación Z

Cuadro 7-1. Algo más sobre las encuestas: errores de muestreo y errores al considerar las muestras.

Volviendo al cuadro 5-3, que trata acerca de sondeos y de la encuesta de Gallup, recordaremos que dejamos sin responder una importante cuestión sobre la letra chica que aparece cerca de los resultados de una encuesta, y que dice algo así como: "Información proveniente de un sondeo telefónico a 1.000 adultos estadounidenses, realizado el 4 y 5 de junio. Error de muestreo ±3%".

Dijimos que una duda común es preguntarse cómo se puede utilizar una cantidad tan pequeña, como 1.000 individuos (aunque rara vez se utiliza una cantidad mucho *menor) para predecir la opinión de todo* el público de los Estados Unidos.

Comencemos con el tema del tamaño de la muestra. De acuerdo a lo aprendido en este capítulo, sabemos que cuando las muestras son de gran tamaño, como lo es una muestra de 1.000 valores, se reduce mucho el desvío estándar de la distribución de medias. Es decir, la distribución de medias muestrales se vuelve muy alta y estrecha, dispersa alrededor de la media poblacional. Por lo tanto, la media de cualquier muestra de ese tamaño está muy cerca de la media poblacional. Para expresarlo de otro modo, la varianza de la distribución de medias, que refleja cuánto tiende a diferir cualquier media muestral de la media poblacional, es la varianza de la población dividida por el tamaño de la muestra. El tamaño de la propia población (de individuos), o la relación del tamaño de la muestra con el de la población, no influye en esta fórmula.

Aun así, nuestra intuición podría continuar diciéndonos que la cantidad necesaria para representar a todo el inmenso público de los Estados Unidos debería ser mayor a sólo 1.000 individuos. Sin embargo, si lo pensamos bien, cuando la muestra es sólo una pequeña parte de una población muy grande, el tamaño absoluto de la muestra es el único determinante de exactitud. Ese tamaño *absoluto determina el impacto de los errores aleatorios* de medición y selección.

Algunas veces sí influye el tamaño relativo de una muestra con respecto a la población; esto ocurre si la población es tan pequeña que, "eliminar" o interrogar a algunos, aumenta las chances de que los restantes sean entrevistados. Pero cuando la población está formada por millones, eliminar a mil o dos mil tendrá un efecto prácticamente nulo en las probabilidades de que sean otros los entrevistados. Una encuesta realizada a 1.000 de entre un millón de votantes, o de entre 10 ó 100 millones de votantes tendrá esencialmente el mismo error casual. Lo importante es reducir *desvíos o errores sistemáticos*, lo cual sólo puede lograrse a través de una planificación muy cuidadosa.

no es diferente al modo usual de convertir una puntuación original en puntuación Z. Sin embargo, debemos ser cuidadosos para no confundirnos, ya que el proceso involucra a más de una media. Es importante recordar que estamos manejando la media muestral como si fuera una simple observación individual. En otras palabras, la fórmula ordinaria (del capítulo 2) para convertir un valor original en puntuación Z es $Z = (X - M)/SD$. En la situación que estamos tratando ahora, en realidad estamos utilizando la siguiente fórmula:

$$Z = \frac{(M - \mu_M)}{\sigma_M} \tag{7-4}$$

Por ejemplo, supongamos que la media muestral es 18 y que la distribución de medias tiene una media de 10 y un desvío estándar de 4. La puntuación Z correspondiente a esta media muestral es +2. Utilizando la fórmula:

$$Z = \frac{(M - \mu_M)}{\sigma_M} = \frac{18 - 10}{4} = \frac{8}{4} = 2$$

La Figura 7-6 ilustra el cálculo anterior.

Ejemplo de prueba de hipótesis con una muestra de más de un individuo

Recordemos el experimento ficticio presentado en los capítulos 1 y 2 acerca de la lectura de oraciones ambiguas. En esos capítulos, simplemente observamos la distribución de los tiempos de lectura cuando las oraciones se presentaban sin contexto. Ahora supondremos que los investigadores quieren probar una teoría sobre la importancia del contexto. Por lo tanto, realizan un estudio analizando los tiempos de lectura cuando existe algún contexto para las oraciones ambiguas

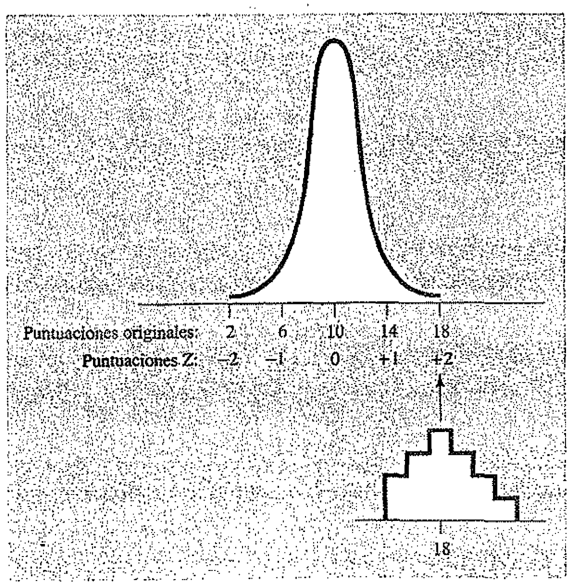


Figura 7-6. Puntuación Z correspondiente a la media de determinada muestra, ubicada en la distribución de medias.

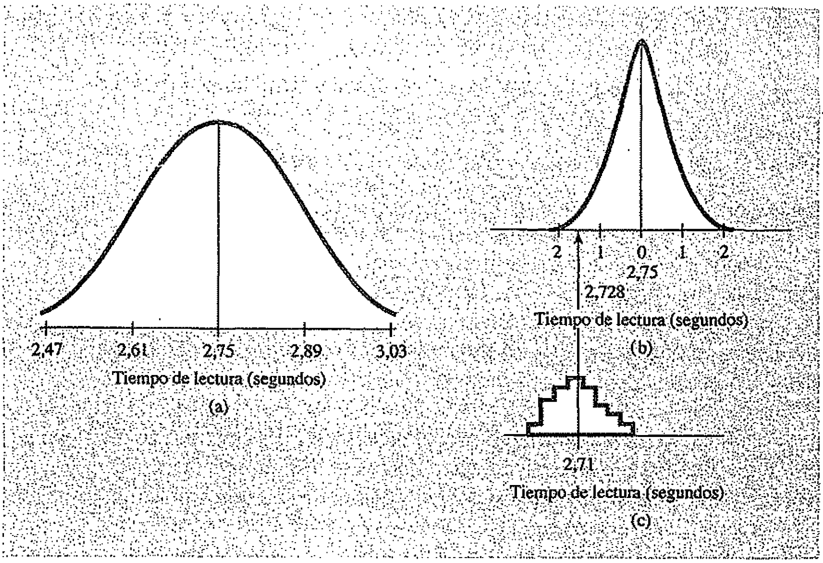


Figura 7-7. Con respecto al experimento ficticio del capítulo 1 acerca de la lectura de oraciones ambiguas, (a) distribución poblacional de valores individuales, (b) distribución de medias y (c) distribución de la muestra.

que hace que sus significados sean un poco más claros. El objetivo es establecer si el tiempo de lectura será más rápido en estas condiciones. Por supuesto, también es posible que al proporcionar un contexto se demore la lectura por el hecho de hacer más complicada la situación.

También supondremos que los investigadores han realizado muchos estudios previos con estas oraciones ambiguas presentadas sin contexto. A partir de esa investigación supondremos que los investigadores confían en que los tiempos de lectura de oraciones ambiguas, sin ningún contexto de la población en general, están distribuidos de forma aproximadamente normal, con una media de 2,75 segundos y una varianza de 0,02 segundos (σ = 0,14 segundos). La figura 7-7a muestra la distribución poblacional a la que nos referimos.

En el estudio que acabamos de describir se prueba a 40 individuos utilizando oraciones ambiguas en contexto. El tiempo medio de lectura es de 2,71 segundos. (En el ejemplo que estamos analizando conocemos la varianza poblacional antes de realizar el estudio. En este tipo de situaciones, la varianza muestral no se utiliza para nada en el proceso de prueba de hipótesis). La figura 7-7c muestra la distribución muestral.[3]

[3] En realidad, este estudio sería mucho mejor si los investigadores tuvieran también otro grupo de participantes a los que se les asignara al azar la realización de una prueba de velocidad de lectura de oraciones ambiguas sin contexto. Confiar en información proveniente de estudios previos es un poco arriesgado, porque las circunstancias en las que se realizaron las pruebas durante uno y otro estudio pueden no ser idénticas. Sin embargo, nos hemos tomado algunas libertades con este ejemplo para ayudarnos a introducir el proceso de prueba de hipótesis de a un paso por vez. En este ejemplo, y en los otros del capítulo, utilizamos situaciones en las que se contrasta una sola muestra con una población "conocida". A partir del capítulo 9, ampliamos el procedimiento de prueba de hipótesis para adaptarlo a situaciones de investigación más realistas, es decir, aquellas que involucran más de un grupo de participantes y que incluyen poblaciones cuyas características se desconocen.

¿Qué deberían concluir los investigadores? Sigamos los pasos de la prueba de hipótesis.

1. **Replantear el problema en función de hipótesis de investigación e hipótesis nula de las poblaciones.** Las dos poblaciones son:

Población 1: participantes que leen oraciones ambiguas en contexto.
Población 2: participantes que leen oraciones ambiguas sin contexto.

La hipótesis de investigación establece que existe una diferencia en el tiempo de lectura entre las dos poblaciones, es decir, que el tiempo de lectura en contexto será diferente al tiempo de lectura sin contexto: $\mu_1 \neq \mu_2$. La hipótesis nula establece que no existe diferencia entre el tiempo de lectura de las dos poblaciones: $\mu_1 = \mu_2$. Cabe mencionar que las hipótesis son no direccionales. Si bien los investigadores esperan que el tiempo de lectura con contexto sea más rápido, no pueden descartar la posibilidad de que el contexto retarde el tiempo de lectura, resultado que además sería bastante interesante.

2. **Determinar las características de la distribución comparativa.** Si la hipótesis nula es verdadera, la población de individuos de la cual proviene nuestra muestra no es diferente de la población 2, cuya media y varianza conocemos. Lo que necesitamos calcular ahora son las características de una distribución de medias de muestras con 40 valores cada una, tomadas de esa población de individuos que conocemos.

Por lo tanto, seguimos las reglas para determinar las características de una distribución de medias: a) la media es igual a la media poblacional, en este caso 2,75 segundos y b) la varianza es igual a la varianza poblacional dividida por la cantidad de valores de cada muestra. Aplicando la fórmula:

$$\sigma_M^2 = \frac{\sigma^2}{N} = \frac{0{,}02}{40} = 0{,}0005$$

El desvío estándar es la raíz cuadrada del resultado anterior, 0,022. Finalmente, c) la forma de la distribución será cercana a una curva normal porque las muestras tienen más de 30 valores cada una. La figura 7-7b ilustra la distribución de medias.

3. **Determinar el punto muestral de corte en la distribución comparativa, a partir del cual debería rechazarse la hipótesis nula.** Supongamos que los investigadores decidieron utilizar un nivel de significación del 5%. Como observamos en el paso 1, han propuesto una hipótesis no direccional, por lo que necesitamos una prueba de dos colas. Acabamos de determinar que la distribución comparativa es normal, por lo tanto, podemos consultar la tabla de áreas bajo la curva normal para encontrar la puntuación Z que marca el 2 1/2% inferior y superior. La tabla nos indica que para rechazar la hipótesis nula a un nivel del 5%, necesitamos una puntuación Z de +1,96 ó mayor, o bien, de −1,96 ó menor. Las dos regiones del 2 1/2%, en las que la hipótesis nula sería rechazada, son las que sostienen a las pequeñas áreas sombreadas (son muy difíciles de ver) en las dos colas de la distribución de medias representada por la figura 7-7b.

4. **Determinar el valor muestral en la distribución comparativa.** La media muestral es de 2,71 (*véase* figura 7-7c). A partir del paso 2, sabemos que la distribución comparativa (nuestra distribución de medias) tiene una media de 2,75 y un desvío estándar de 0,022. Aplicando la fórmula:

$$Z = \frac{(M - \mu_M)}{\sigma_M} = \frac{2{,}71 - 2{,}75}{0{,}022} = \frac{-0{,}04}{0{,}022} = -1{,}82$$

5. **Comparar los valores de los pasos 3 y 4 para decidir si se rechaza o no la hipótesis nula.** La puntuación Z necesaria para rechazar la hipótesis nula es ±1,96. La puntuación Z que obtuvimos es de sólo −1,82. Por lo tanto, no podemos rechazar la hipótesis nula: el experimento no es concluyente. Podemos observar el resultado gráficamente en la figura 7-7b, la cual muestra que la ubicación de la media de nuestra muestra en la distribución de medias no es tan extrema como para que sea claramente inverosímil que pueda ser seleccionada de esa distribución.

No obstante, el resultado es casi lo suficientemente extremo como para rechazar la hipótesis nula. Por lo tanto, los investigadores podrían indicar que el resultado fue "casi significativo" o "cercano a la significación", agregando tal vez que "$p<0,10$". (El punto de corte para la significación a un nivel de 0,10, en una prueba de dos colas, es de ±1,64). Pero con un resultado límite como este, el mejor consejo es repetir el experimento, tal vez con más participantes. (El capítulo 8 incluye una exposición acerca de los efectos de aumentar la cantidad de participantes con respecto a la probabilidad de que el experimento produzca un resultado significativo).

Otro ejemplo de prueba de hipótesis con una muestra de más de un individuo

Este es otro ejemplo ficticio. Dos psicólogos especializados en educación están estudiando los efectos que tienen las instrucciones en las pruebas de nivel académico cronometradas. Su teoría establece que si se instruye a quienes van a realizar la prueba para que contesten cada pregunta con la primera respuesta que les venga a la mente, los resultados de las pruebas serán mejores.

Para analizar esa teoría, los investigadores organizaron que 64 alumnos de quinto grado, seleccionados en forma aleatoria, rindieran una prueba de nivel académico estándar. La prueba se toma de la manera comúnmente utilizada, con una sola excepción. Como parte del estudio, las instrucciones para la prueba incluyen una indicación adicional que aconseja a los alumnos a responder cada pregunta con la primera respuesta que les venga a la mente. Cuando se toma la prueba en la forma acostumbrada (es decir, sin esa indicación extra en las instrucciones) la media es de 200, el desvío estándar de 48 y la distribución, representada gráficamente en la figura 7-8a, aproximadamente normal.

¿Qué tipo de resultado necesitarían los psicólogos expertos en educación para concluir que el procedimiento tiene algún efecto?

1. **Replantear el problema en función de hipótesis de investigación e hipótesis nula de las poblaciones.** Las dos poblaciones son:

Población 1: alumnos de quinto grado que reciben instrucciones especiales.
Población 2: alumnos de quinto grado que no reciben instrucciones especiales.

La hipótesis de investigación establece que la población de alumnos de quinto grado que rinden la prueba con las instrucciones especiales, obtendrá puntuaciones más altas que la población de alumnos que rinden la prueba en la forma acostumbrada; $\mu_1 > \mu_2$. La hipótesis nula establece que las puntuaciones de la población 1 no serán mayores que las de la población 2; $\mu_1 \leq \mu_2$. (Cabe destacar que las hipótesis son direccionales).

2. **Determinar las características de la distribución comparativa.** A partir del estudio obtenemos la media de una muestra formada por 64 observaciones (en este caso, de alumnos de quinto grado). La distribución comparativa debe ser la distribución de medias de muestras formadas por 64 valores cada una. Esa distribución tendrá una media de 200 (igual a la media poblacional) y la varianza será igual a la varianza poblacional dividida por la cantidad de individuos en la muestra. La varianza poblacional es 2.304 (el desvío estándar poblacional de 48 elevado al cuadrado) y el tamaño de la muestra es 64. Por lo tanto, la varianza de la distribución de medias será

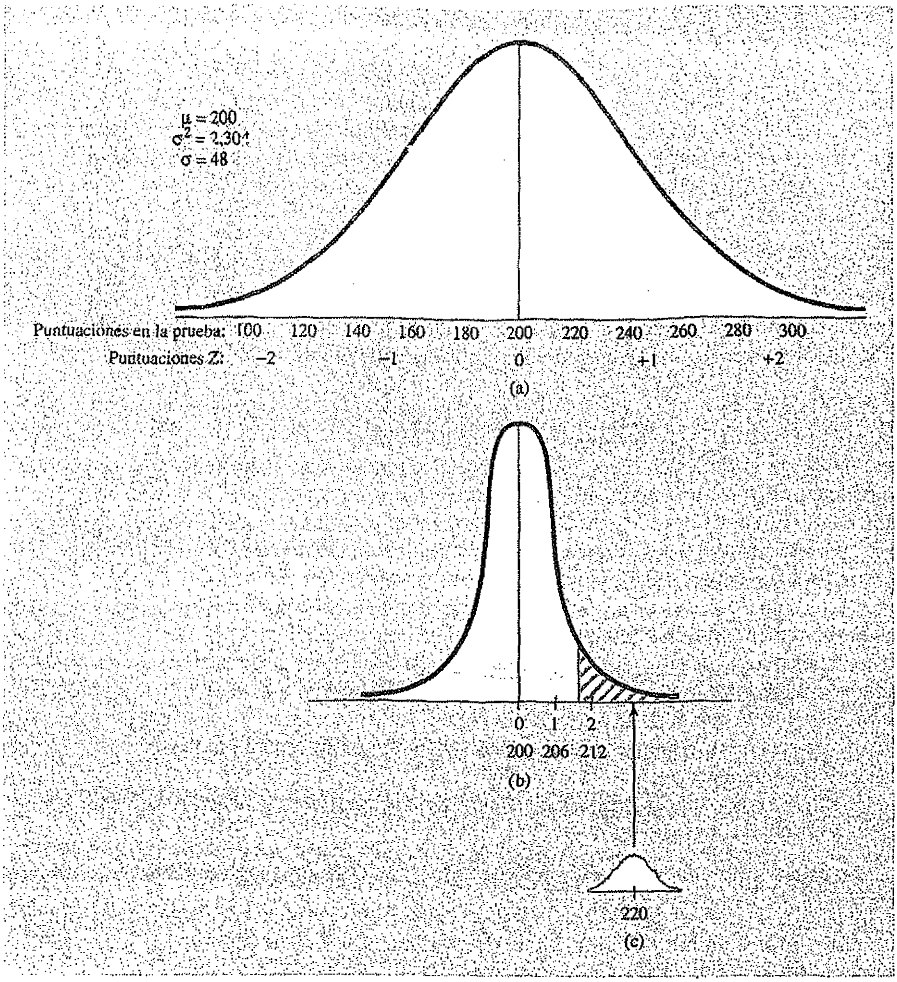


Figura 7-8. Con respecto al estudio ficticio basado en el desempeño en una prueba estándar de nivel académico, (a) distribución poblacional de valores individuales, (b) distribución de medias (distribución comparativa) y (c) distribución de la muestra.

2.304/64, es decir, 36. El desvío estándar de la distribución de medias es la raíz cuadrada de 36, o sea, 6. Finalmente, dado que en la muestra hay más de 30 individuos, la forma de la distribución de medias será aproximadamente normal. La figura 7-8b muestra la distribución de medias que acabamos de describir.

3. ***Determinar el punto muestral de corte en la distribución comparativa, a partir del cual debería rechazarse la hipótesis nula.*** Una vez más, supongamos que los investigadores adoptan el nivel de significación usual del 5%. Los investigadores que realizan este estudio tienen

una predicción claramente direccional, y realmente no están interesados en ningún efecto en dirección contraria. (Si las instrucciones especiales no mejoran las puntuaciones de la prueba, no serán utilizadas en el futuro. Cualquier posible resultado que muestre un efecto negativo es irrelevante). Por lo tanto, los investigadores rechazarán la hipótesis nula si el resultado se encuentra dentro del 5% superior de la distribución comparativa. La distribución comparativa (la distribución de medias) es una distribución normal, por ende, podemos determinar el 5% superior a través de la tabla de áreas bajo la curva normal. La parte que nos interesa bajo la curva normal comienza en una puntuación Z de +1,64, y el área sombreada de la figura 7-8b muestra ese 5% superior.

4. Determinar el valor muestral en la distribución comparativa. Los 64 alumnos de quinto grado que realizaron la prueba aplicando las instrucciones especiales tenían una puntuación media de 220. (La figura 7-8c grafica la distribución de esa muestra). Una media de 220 se encuentra a 3,33 desvíos estándar por encima de la media de la distribución de medias:

$$Z = \frac{(M - \mu_M)}{\sigma_M} = \frac{220 - 200}{6} = \frac{20}{6} = 3,33$$

5. **Comparar los valores de los pasos 3 y 4 para decidir si se rechaza o no la hipótesis nula.** Establecimos que la puntuación Z mínima necesaria para rechazar la hipótesis nula es +1,64. La puntuación Z correspondiente a la media muestral es de +3,33. Por lo tanto, los psicólogos expertos en educación pueden rechazar la hipótesis nula y concluir que se sostiene la hipótesis de investigación. Para decirlo con otras palabras, el resultado es estadísticamente significativo al nivel $p<0,05$. El resultado puede verse reflejado en la figura 7-8b, donde se observa cuán extrema es la media muestral en la distribución de medias (la distribución que se aplicaría si la hipótesis nula fuera verdadera). La conclusión final es que entre alumnos de quinto grado como los analizados, las instrucciones especiales sin duda mejoran las puntuaciones en las pruebas.

ESTIMACIÓN E INTERVALOS DE CONFIANZA

La prueba de hipótesis es el tema central de este libro. Sin embargo, existe otro tipo de cuestión estadística relacionada con la distribución de medias que, algunas veces, resulta importante para la psicología. Esa otra cuestión es la estimación de la media de una población desconocida sobre la base de los valores muestrales. El tema es importante, por ejemplo, en investigaciones a través de encuestas, y como veremos más adelante, también puede ser importante como método alternativo de la prueba de hipótesis.

Estimaciones puntuales y estimaciones por intervalos

La mejor estimación de la media poblacional es la media muestral. En el estudio realizado con los alumnos de quinto grado que recibieron instrucciones especiales, la puntuación media de la muestra formada por los 64 individuos analizados era de 220. Por lo tanto, 220 es la mejor estimación de la media correspondiente a la población desconocida de alumnos de quinto grado que podrían recibir instrucciones especiales. En este caso, estamos estimando el valor específico de la media poblacional. Cuando estimamos el valor específico de un parámetro poblacional, hablamos de una **estimación puntual**.

También podemos encontrar un intervalo de posibles medias verosímil, el cual incluya la media poblacional. Por ejemplo, podríamos decir estimativamente que un intervalo de 200 a 240 incluye la verdadera media poblacional de alumnos de quinto grado que reciben instrucciones especiales.[4] A esto se lo denomina **estimación por intervalos**.

Principio general y terminología relacionada con los intervalos de confianza

Cuanto más amplia sea la estimación por intervalos, mayor será la certeza de que incluya la verdadera media poblacional. En el ejemplo de los alumnos de quinto grado podríamos estar bastante seguros de que el intervalo de 100 a 340 incluye la verdadera media poblacional. Pero estaríamos arriesgándonos al fracaso si estimáramos que la verdadera media poblacional está incluida en el intervalo de 219 a 221.

En general, necesitamos un intervalo lo suficientemente amplio como para asegurarnos que incluya la media poblacional. A esto se lo denomina **intervalo de confianza** (a veces se abrevia IC). Si queremos estar 95% seguros, necesitamos un **intervalo del 95% de confianza**. Un intervalo del 95% de confianza, en el ejemplo de los alumnos de quinto grado, abarca desde 208,24 hasta 231,76. Es decir, sobre la base de la muestra analizada, podemos estar un 95% seguros de que un intervalo de 208,24 a 231,76 incluye la verdadera media poblacional. (Pronto aprenderemos a calcular los intervalos). Los límites superiores e inferiores de los intervalos de confianza se denominan límites de confianza. En el ejemplo que estamos analizando, los **límites de confianza** son 208,24 y 231,76.

Si queremos tener una seguridad aún mayor al 95%, necesitamos un intervalo más amplio. En nuestro ejemplo, los límites de confianza de un **intervalo del 99% de confianza** son 204,58 y 235,42.

Determinación de los límites de confianza

Los límites de confianza se basan en la distribución de medias. Lo que necesitamos saber es dónde comienza y termina el 95% central de las medias en esa distribución. Por lo tanto, necesitamos encontrar los puntos de corte correspondientes al 2,5% inferior y al 2,5% superior, lo que deja un total del 95% en el centro. (Para los intervalos del 99% de confianza necesitaríamos calcular los puntos que marcan el 0,5% superior e inferior, con lo cual dejaríamos un 99% en el centro).

Comencemos con el límite inferior. Como siempre, lo más sencillo es pensar en función de las puntuaciones Z. La puntuación Z que marca el 2,5% inferior en una curva normal es –1,96. (*Este dato lo encontraremos en la tabla de áreas bajo la curva normal). El ejemplo tiene una media* de 220 y un desvío estándar de la distribución de medias igual a 6. Por lo tanto, en esta distribución de medias, una puntuación Z de –1,96 corresponde a 208,24. (Es decir, utilizando el procedimiento usual para convertir una puntuación Z en una puntuación original, convertimos la puntuación Z –1,96 en la puntuación original 208,24).

El cálculo del límite superior funciona de la misma manera. La puntuación Z que marca el 2,5% superior es +1,96 que, en la distribución de medias, equivale a 231,76.

Pasos a seguir para el cálculo de los intervalos de confianza

A continuación presentamos los tres pasos para calcular intervalos de confianza. Al seguir estos pasos se supone que la distribución de medias es una distribución aproximadamente normal.

1. Determinar las características de la distribución de medias utilizando el cálculo acostumbrado. No obstante, cabe destacar que estamos interesados en la distribución de medias co-

[4] Según la lógica matemática de la estadística inferencial, debemos considerar la media poblacional como algo fijo. Los intervalos de confianza pueden variar, pero la media poblacional es fija. Por lo tanto, podemos decir que estamos 95% seguros de que nuestro intervalo de confianza incluye la media poblacional. No deberíamos decir que las chances de que la media poblacional se encuentre dentro del intervalo de confianza son del 95%.

rrespondiente a la población que representa la muestra que estamos analizando (lo que hemos llamado población 1), y no en la distribución de medias correspondiente a la población con la cual la estamos comparando (población 2). Se estima entonces que la media de la distribución de medias es la media muestral. En cuanto a la varianza, afortunadamente, por lo general suponemos que la varianza de las dos poblaciones es la misma. Consecuentemente, podemos utilizar la varianza conocida de la población dada (población 2) como base para calcular la varianza de la distribución de medias de la población en la que estamos interesados (población 1). (La varianza de la distribución de medias se basa sólo en la varianza de la población y en el tamaño de la muestra. Por lo tanto, la varianza de la distribución de medias será igual para ambas poblaciones).

2. Utilizar la tabla de áreas bajo la curva normal para encontrar las puntuaciones Z que coinciden con los porcentajes superiores e inferiores que nos interesan. Para un intervalo del 95% de confianza, debemos buscar la puntuación Z que coincide con el 2,5% inferior y el 2,5% superior. Para un intervalo del 99% de confianza, debemos buscar la puntuación Z que coincide con el 0,5% inferior y el 0,5% superior.

3. Convertir las puntuaciones Z en puntuaciones originales de la distribución de medias. Esos son los límites de confianza superior e inferior.

Otro ejemplo de cálculo del intervalo de confianza

Analicemos otro ejemplo. Calculemos el intervalo de confianza para el estudio basado en las oraciones ambiguas en el caso de los participantes que las leían en contexto. En ese caso, los 40 individuos analizados de la manera mencionada tenían un tiempo medio de lectura de 2,71 segundos, y sabíamos por investigaciones anteriores que la población de individuos que leía oraciones ambiguas sin ningún contexto tenía una varianza de 0,02 segundos. Con esta información estamos preparados para calcular el intervalo de confianza.

1. Determinar las características de la distribución de medias. La media será de 2,71 segundos. Suponemos que la población de individuos puestos a prueba leyendo oraciones ambiguas en contexto tendrá la misma forma y varianza que la población que lee sin contexto ($\sigma^2 = 0,02$). Por lo tanto, la distribución de medias será normal y tendrá una varianza igual a 0,02/40 ó 0,0005. El desvío estándar es la raíz cuadrada de este resultado, es decir, 0,022. (Cabe mencionar que obtuvimos el mismo desvío estándar de la distribución de medias que calculamos anteriormente al realizar la prueba de hipótesis y concentrarnos en la distribución de medias para la población que lee las oraciones sin contexto).

2. Utilizar la tabla de áreas bajo la curva normal para encontrar las puntuaciones Z que corresponden al porcentaje superior e inferior que hayamos elegido. Suponiendo que queremos establecer el intervalo usual del 95% de confianza, entonces buscaremos las puntuaciones Z que corresponden al 2,5% superior e inferior. Como vimos anteriormente, el resultado es ±1,96.

3. Convertir esas puntuaciones Z en puntuaciones originales de la distribución de medias. Con una media de 2,71 y un desvío estándar de 0,022, una puntuación Z de –1,96 es igual a una puntuación original de 2,71 – (0,022 x 1,96), que da como resultado 2,667. De modo similar, una puntuación Z de +1,96 es igual a una puntuación original de 2,71 + (0,022 x 1,96), que da como resultado 2,753. Por lo tanto, los límites del 95% de confianza son 2,667 a 2,753. Es decir, sobre la base de los resultados del estudio, tenemos un 95% de confianza de que la verdadera media del tiempo de lectura de oraciones ambiguas presentadas en contexto se encuentra entre 2,667 y 2,753 segundos.

La sutil lógica de los intervalos de confianza

La lógica de los intervalos de confianza es un poco más sutil de lo que podría parecer a simple vista. Esa sutileza está relacionada con el hecho de que los intervalos de confianza son estimaciones basadas sólo en información sobre una muestra. Es decir, tal como sucede con la prueba de hipótesis, los intervalos de confianza involucran inferencia estadística acerca de una población sobre la base de datos obtenidos de una muestra.

Para aclarar los puntos más delicados de esta lógica, es útil imaginar que, de algún modo, conocemos la media real de la población. Por ejemplo, supongamos que de alguna forma sabemos que la población de alumnos de quinto grado que recibe instrucciones especiales (la población 1 en los ejemplos anteriores) tiene una media de 210. (Nuestra atención está puesta ahora en la población acerca de la cual estamos realizando estimaciones basándonos en la muestra. No debemos confundir esa población con la población 2, la que conocíamos desde el principio y en la cual los alumnos de quinto grado no recibían instrucciones especiales). En tal población, un intervalo del 95% de confianza estaría ubicado alrededor de su media de 210. Utilizando el procedimiento de cálculo que conocemos, podemos calcular que existe una probabilidad del 95% de que cualquier media muestral se encuentre entre 198,24 y 221,76 (*véase* figura 7-9, intervalo a).

Si de hecho la verdadera media de la población fuera 210, no hubiera sido sorprendente que, al analizar una muestra de 64 alumnos, los investigadores especializados en educación obtuvieran una media de 220. Esta media muestral estaría perfectamente incluida en los límites del 95% de la distribución de medias.

Sin embargo, en las investigaciones, en general no conocemos la media de la población que estamos analizando. En realidad, los investigadores especializados en educación no tendrían forma de saber que la verdadera media de la población de alumnos de quinto grado que reciben instrucciones especiales es 210. Todo lo que conocen es la media de su muestra particular. Aun así, pueden utilizar la media muestral como estimación de la media poblacional. Basándose en esa estimación, pueden calcular un intervalo del 95% de confianza, que ya hemos determinado que abarcaría desde 208,24 a 231,76. Con bastante seguridad, este intervalo de confianza contendría la media real de la población, por lo que la confianza en el intervalo está justificada. (*Véase* figura 7-9, intervalo b)

De todos modos, supongamos que los investigadores especializados en educación hubieran realizado el estudio y hubieran descubierto que su muestra tenía una media de 190. Continuamos suponiendo que la verdadera media de la población (que los investigadores desconocen) es 210, y que el 95% de las veces las muestras de 64 individuos tomadas de esta población deberían ubicarse entre 198,24 y 221,76. Por lo tanto, obtener una muestra con una media de 190 es bastante improbable; en verdad la probabilidad es menor al 5%, pero posible. De hecho, esperamos que el 5% de las veces las muestras presenten medias fuera del intervalo del 95%.

Supongamos que los psicólogos especializados en educación continúan con el estudio y calculan el intervalo de confianza utilizando la media muestral de 190 como estimación de la media poblacional. Siguiendo las reglas usuales de cálculo, obtendrán un intervalo del 95% de confianza que abarque desde 178,24 hasta 201,76. Por lo tanto, calcularían un intervalo de confianza que no incluye la verdadera media poblacional. (*Véase* figura 7-9 intervalo c).

En resumen, cuando la media estimada se encuentra dentro de los límites del 95% de la verdadera media poblacional, el intervalo de confianza incluirá la verdadera media. Afortunadamente, el 95% de las veces la media estimada se encuentra dentro de los límites del 95% de la media poblacional real. En esos términos, un 5% de las veces el intervalo de confianza calculado sobre la base de la media estimada no incluirá la verdadera media.

En otras palabras. El 95% de las veces que calculemos un intervalo de confianza, éste incluirá la verdadera media; y el 5% de las veces, no. Por eso decimos que estamos un 95% seguros de que el intervalo incluye la verdadera media. Sin embargo, nunca sabremos con certeza si nos encontramos en la situación del 95% ó del 5%. Siempre existe un 5% de chances de que la verdadera media no esté incluida para nada dentro de la sección calculada.

Intervalos de confianza y prueba de hipótesis

Además de su valor en cuanto a la estimación de la media poblacional, también podemos utilizar los intervalos de confianza para realizar pruebas de hipótesis. Si un intervalo de confianza no incluye la media de la distribución de la hipótesis nula, entonces el resultado es significativo. Esto se debe a que estamos un 95% seguros de que el intervalo incluye la verdadera media poblacional, y si ese intervalo del 95% no incluye la media de la población 2, entonces existe menos de un 5% de chances de que la muestra hubiera podido surgir de la población 2. En el ejemplo de las instrucciones especiales para rendir el examen de nivel, el intervalo del 95% de confianza que abarcaba desde 208,24 hasta 231,76 no incluye la media de 200 correspondiente a la población de alumnos de quinto grado que rinden el examen sin las instrucciones especiales. El resultado que acabamos de mencionar es coherente con la conclusión que sacamos con anterioridad en este mismo capítulo, según la cual el resultado era significativo utilizando el nivel 0,05. En el ejemplo de las oraciones ambiguas, el intervalo del 95% de confianza con respecto a aquellos que leían las oraciones en contexto, abarcaba desde 2,667 a 2,753 segundos. Ese intervalo en efecto, incluye la media de tiempo de lectura (2,75) de la población que leía las oraciones sin contexto. Por lo tanto, tal como concluimos cuando utilizamos el procedimiento de prueba de hipótesis, el resultado no es significativo aplicando el nivel 0,05.

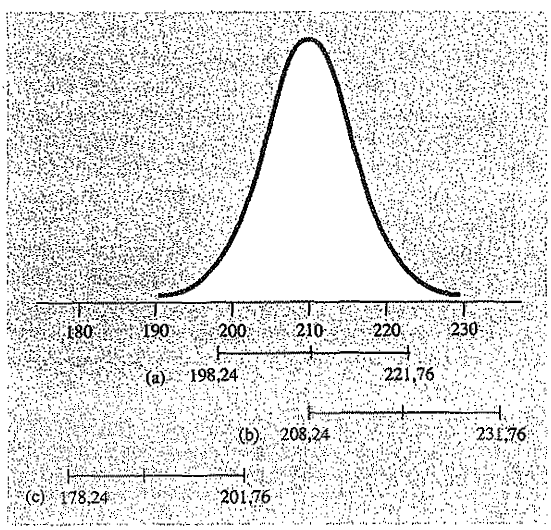


Figura 7-9. Ejemplos de intervalos del 95% de confianza comparados con distribuciones de medias basadas en (a) una media poblacional conocida, igual a 210; (b) una media muestral igual a 220, y (c) una media muestral de 190.

CONTROVERSIAS Y LIMITACIONES: ¿INTERVALOS DE CONFIANZA O PRUEBAS DE SIGNIFICACIÓN?

El alumno recordará que en el capítulo 6 mencionamos que, en la actualidad, existe un enérgico debate entre los psicólogos acerca de la prueba de significación. Entre los principales temas de debate se ha propuesto que los psicólogos utilicen los intervalos de confianza en lugar de las pruebas de significación.

Aquellos que están a favor de reemplazar las pruebas de significación con los intervalos de confianza (p. ej. Cohen, 1994; Hunter, 1997; Schmidt, 1996) citan varias ventajas importantes. Primero, como observamos anteriormente, los intervalos de confianza contienen toda la información clave de una prueba de significación,[5] pero además proporcionan información adicional: la estimación del intervalo de valores dentro del cual podemos estar bastante seguros de que se encuentra la verdadera media poblacional. Una segunda ventaja es que concentran la atención en la estimación y no en la prueba de hipótesis. Algunos investigadores argumentan que el objetivo de la ciencia es proporcionar estimaciones numéricas de efectos, no sólo decisiones en cuanto a si un efecto es diferente de cero. Es decir, con las estimaciones (puntuales y por intervalos), tenemos una idea clara del grado de importancia del efecto y del nivel de precisión de la estimación. Con las pruebas de hipótesis, sabemos si el efecto puede suceder en la dirección predicha, pero no el grado de importancia del efecto en esa dirección.

Los intervalos de confianza son particularmente valiosos cuando los resultados no son significativos (Frick, 1995), porque conocer el intervalo de confianza otorga una idea de cuán lejos de la ausencia de efecto es probable encontrar la verdadera media. Si todo el intervalo de confianza se encuentra cerca de la ausencia de efecto, podemos tener la certeza de que si aún existe algún efecto verdadero, éste probablemente sea pequeño. Por ejemplo, supongamos que se estudia un grupo de personas después de que son expuestas a un procedimiento que pretende afectar el *IC*. La media del grupo es 102, y el intervalo de confianza abarca desde 99 hasta 105. Esto daría un resultado no significativo porque el intervalo incluye el valor 100, que es el *IC* medio de la población que no recibe el procedimiento especial. Al mismo tiempo, dado que el intervalo de confianza incluye otros números diferentes de 100, en realidad es posible que exista un efecto real. Sin embargo, el punto clave es que si de hecho existiera un efecto real, es probable que sea muy pequeño, ya que estamos un 95% seguros de que ese efecto no implicaría más que una disminución de un punto o un aumento de 5 puntos. Por otro lado, supongamos que el intervalo de confianza para este mismo estudio era de 89 a 115. Este resultado también sería no significativo (porque incluye el valor 100). Sin embargo, nos indicaría que el estudio es realmente no concluyente: es posible que haya muy poco o ningún efecto (que la media poblacional de aquellos que reciben el procedimiento sea cercana a 100), pero también es posible que exista un efecto substancial (que la verdadera media poblacional de aquellos que reciben el procedimiento implique una disminución de hasta 11 puntos de *IC*, o un aumento de hasta 15 puntos de *IC*).

Una tercera ventaja, sostenida por aquellos que proponen los intervalos de confianza para reemplazar las pruebas de significación, es que existe menos probabilidad de que los investigadores los utilicen erróneamente. Como observamos en el capítulo 6, un error generalizado en la utilización de las pruebas de significación es concluir que un resultado no significativo implica que

[5] Algunos de los que proponen los intervalos de confianza para reemplazar la prueba de significación sostienen que deberíamos ignorar el vínculo con la prueba de hipótesis. Esta es la posición más radical en contra de la prueba de significación. Es decir, estos psicólogos argumentan que todo el enfoque debería concentrarse en la estimación, y que la prueba de significación de cualquier tipo debería ser irrelevante. En el capítulo 8, veremos los fundamentos de esta posición, junto con los argumentos contrarios.

no existe ningún efecto. Con los intervalos de confianza es más difícil caer en este tipo de error. Si bien el intervalo de confianza que arroja un resultado no significativo incluirá la media esperada correspondiente a la ausencia de efecto, también incluirá otros valores posibles. Así, nos recuerda que la verdadera media poblacional podría muy bien ser diferente de la media correspondiente a la ausencia de efecto.

A pesar de estas aparentes ventajas, es extremadamente raro encontrar intervalos de confianza en la mayoría de los diferentes tipos de publicaciones científicas psicológicas. En parte, esto probablemente se debe a la tradición y a que la mayoría de los psicólogos han sido capacitados para utilizar las pruebas de significación, por lo que están mucho más acostumbrados a ellas. En una publicación científica, los intervalos de confianza también requieren una mayor descripción. Por ejemplo, qué sucedería en el caso de que tuviéramos una tabla de resultados más amplia. Sería sencillo agregar un asterisco en cada número para mostrar su significación, por lo cual una tabla diseñada de ese modo es fácil de leer. Con los intervalos de confianza, en lugar de un asterisco, necesitaríamos dos números extra para cada resultado (los límites de confianza superior e inferior).

Otros psicólogos (p. ej. Abelson, 1997; Harris, 1997) indican dos razones para no abandonar por completo las pruebas de significación a favor de los intervalos de confianza. Primero, en algunos procedimientos estadísticos avanzados es posible realizar pruebas de significación, pero no es posible calcular intervalos de confianza. Segundo, del mismo modo que es posible cometer errores con las pruebas de significación, también es posible cometer otros tipos de errores con los intervalos de confianza, especialmente debido a que la mayoría de los psicólogos que realizan investigaciones tienen menos experiencia en la utilización de estos últimos.

Finalmente, la cuestión de los intervalos de confianza, en contraposición con la significación, tiene sus raíces en una mayor controversia entre estimación y prueba de hipótesis, controversia que trataremos en el capítulo 8. Sin embargo, para anticipar esa exposición, podemos señalar aquí que los intervalos de confianza, por lo general, tienen mucho más sentido en situaciones de investigación aplicada, mientras que las pruebas de significación, con frecuencia, tienen mucho más sentido en investigaciones con una orientación más teórica.

Cualquiera sea el resultado de esta controversia sobre intervalos de confianza, es importante comprenderlos, ya que podremos encontrarlos ocasionalmente al leer material relacionado con la investigación, y es posible que en el futuro aparezcan con más asiduidad. No obstante, en la actualidad no aparecen con frecuencia. Por eso, y para que la cantidad de material a aprender sea manejable, decidimos no hacer hincapié en el tema de los intervalos de confianza en los próximos capítulos de este libro que tratan principalmente sobre pruebas de significación en distintos tipos de investigaciones.

DESVÍO ESTÁNDAR DE LA DISTRIBUCIÓN DE MEDIAS MUESTRALES, PRUEBAS DE HIPÓTESIS SOBRE MEDIAS E INTERVALOS DE CONFIANZA SEGÚN SE DESCRIBEN EN PUBLICACIONES CIENTÍFICAS

Como hemos mencionado varias veces, es bastante raro en psicología realizar investigaciones en las que se conoce la media y el desvío estándar de la población. Hemos enseñado ese tipo de situación principalmente porque es la base para comprender la prueba de hipótesis en situaciones comunes de investigación. En los raros casos en los que se realiza una investigación en la que se conoce la distribución poblacional, con frecuencia se describe a través de una **prueba Z**, porque es la puntuación Z la que se compara con la distribución normal.

Analicemos un ejemplo. Como parte de un estudio más amplio, Wiseman (1997) tomó una prueba que mide el nivel de soledad entre un grupo de alumnos universitarios de Israel. Como primer paso para analizar los resultados, Wiseman controló que la media de la prueba del nivel de soledad no fuera diferente a la de una distribución de población conocida a través de un gran estudio norteamericano de alumnos universitarios, realizado con anterioridad por Russell et al. (1980). Así, Wiseman informó:

> El valor medio de soledad de la muestra israelí actual era similar a los de la muestra universitaria realizada por Russell et al. (1980), tanto para hombres (Israelí: M = 38,74, SD = 9,30; Russell: M = 37,06, SD = 10,91; z = 1,09, NS) como para mujeres (Israelí: M = 36,39, SD = 8,87; Russell: M = 36,06, SD = 10,11; z = 0,25, NS) (p. 291).

En este ejemplo, el investigador nos da el desvío estándar tanto de la muestra que está analizando (el grupo israelí) como de la población (la información tomada del estudio de Russell). Sin embargo, al seguir los pasos para calcular cada Z (el valor muestral en la distribución de medias), sólo habría utilizado el desvío estándar poblacional. Observemos también que el investigador consideró la falta de significación de la diferencia como sustento para determinar que las medias muestrales eran "similares" a las medias poblacionales. De todos modos, el investigador fue muy cuidadoso en no pretender que estos resultados mostraban que no existía diferencia alguna.

De todos los temas que hemos tratado en éste capítulo, el que más frecuentemente se menciona en una publicación científica es el desvío estándar de la distribución de medias, utilizado para indicar la cantidad de variación que podría esperarse entre las medias de muestras de determinado tamaño tomadas de la población. En ese contexto, se lo denomina comúnmente **error estándar,** abreviado *SE*. Por ejemplo, Foertsch y Gernsbacher (1997) realizaron un estudio para analizar el efecto de la utilización del pronombre **ellos** para evitar determinar el sexo de la persona a la cual se hace referencia, aunque tradicionalmente este uso se considera gramaticalmente incorrecto. Foertsch y Gernsbacher elaboraron la hipótesis de que utilizar **ellos** del modo mencionado no tendría un efecto importante en el tiempo de lectura. Analicemos la siguiente oración: "Un conductor de camiones nunca debería conducir cansado, aun cuando ella esté intentando realizar una entrega a tiempo, porque muchos accidentes son causados por conductores que se duermen sobre el volante". Como parte del estudio, los investigadores midieron el tiempo de lectura de esta versión de la oración y del de otras dos versiones, una reemplazando **ella** por él y otra remplazando **ella** por **ellos**. En esta oración, el antecedente (la primera cláusula) se refiere a un conductor de camiones, una profesión típicamente masculina. En otras oraciones utilizadas por los investigadores, los antecedentes eran típicamente femeninos (una enfermera) o neutrales (un corredor). Estos son algunos de los resultados:

> Tratándose de antecedentes masculinos, las cláusulas con el pronombre **ella** (M = 59,5, SE = 2,05) se leyeron significativamente más despacio que las cláusulas con el pronombre **él** (M = 54,8, SE = 1,77) o las cláusulas con el pronombre **ellos** (M = 55,3, SE = 1,77) [...] Tratándose de antecedentes femeninos, las cláusulas con el pronombre **él** (M = 58,7, SE = 1,66) se leyeron significativamente más despacio que las cláusulas con el pronombre **ella** (M = 52,9, SE = 1,64) o las cláusulas con el pronombre **ellos** (M = 52,7, SE = 1,67)". (p. 108)

El informe anterior nos proporciona el patrón de medias y una idea clara de la exactitud de esas medias como estimaciones de las medias poblacionales. Analicemos las consecuencias del primer error estándar (2,05). Conocer este dato nos indica que la media del tiempo de lectura de oraciones con antecedentes masculinos y cláusulas con el pronombre **ella,** es de más de 2 errores estándar por encima del tiempo de lectura de oraciones con cláusulas con los pronombres **él** o **ellos**.

Cuando los investigadores informan el error estándar de un resultado, también proporcionan información para calcular el intervalo de confianza. Por ejemplo, suponiendo la distribución normal, calculemos el intervalo del 95% de confianza para oraciones con antecedentes masculinos y cláusulas con el pronombre **ella**. Dado que el *SE* (que es otra manera de denominar al desvío estándar de la distribución de medias) es 2,05, el límite superior del 95% de confianza es la media más el resultado de 1,96 x 2,05. Es decir, 59,5 más 1,96 x 2,05, lo que da 63,52. El límite inferior es 55,48. Por lo tanto, tenemos un 95% de confianza de que el intervalo de 55,48 a 63,52 incluye la verdadera media poblacional.

Los errores estándar aparecen también con frecuencia en publicaciones científicas representados por segmentos ubicados encima de las barras de un diagrama de barras. Esos segmentos sobre las barras principales también se denominan **barras de errores estándar**. Por ejemplo, la figura 7-10, tomada de la publicación de Foertsch y Gernsbacher, grafica los mismos resultados explicados anteriormente (más otros adicionales).

Como ya hemos observado, los intervalos de confianza rara vez se mencionan directamente en publicaciones científicas psicológicas, aunque pueden ser más frecuentes en el futuro. El siguiente es un ejemplo que pudimos encontrar entre material de lectura actual. Chiu, Hong y Dweck (1997) realizaron un estudio que se concentraba en la tendencia de algunos individuos a creer que las características de las personas son permanentes; Chiu et al. clasificaron a estos individuos como "teóricos de entidades", porque ven a las otras personas como entidades invariables.

En particular, los investigadores intentaban probar la teoría de que estos teóricos tendrían más propensión a tomar un sólo hecho como evidencia de que la persona presenta determinada característica permanente. Como parte del estudio, describieron a los participantes una situación

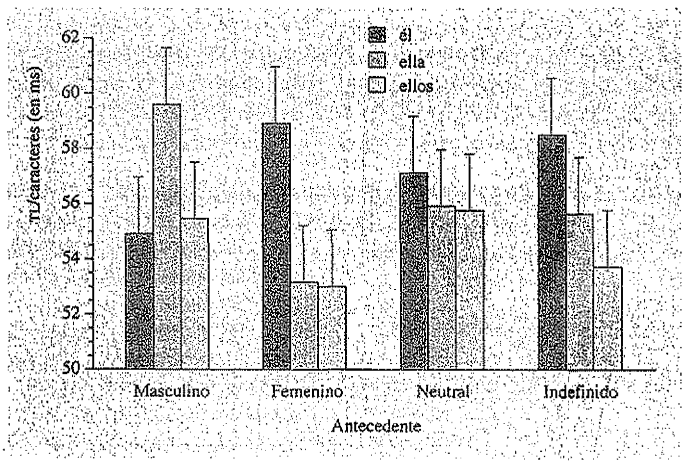


Figura 7-10. Efectos del tipo de antecedente (masculino, femenino, neutro o indefinido) y del pronombre (él, ella o ellos) en TL (Tiempo de lectura por caracter), cuando las oraciones se utilizaron sin referencia. (Experimento 1). [**Fuente**: Foertsch, J., & Gernsbacher, M. A. (1997), fig. 1. "En busca de la neutralidad del género: ¿Es el "ellos" singular un sustituto cognitivamente eficiente del "él" genérico?", *Ciencia Psicológica* [*Psychological Science*], 8, 108. Copyright, 1997, por la Sociedad Americana de Psicología [American Psychological Society]. Reimpreso con autorización.]

en la que una persona se comportaba de forma más amistosa que otra, y luego les preguntaron qué persona sería más propensa a ser amistosa en el futuro.

Así informaron Chiu et al. uno de sus descubrimientos acerca de los teóricos de entidades: "Para ellos, si una persona resultó ser más amistosa que otra en una determinada situación, es más probable que la misma relación se generalice a otras situaciones totalmente diferentes" (p. 23). El sustento estadístico de esta conclusión fue descripto de la siguiente manera: "La predicción global de los 'teóricos de entidades' [acerca de la probabilidad de que la persona fuera amistosa] fue significativamente mayor a 0,50 (95% IC = 0,5583 ± 0,0348)" (p. 23). Es decir que podemos tener un 95% de confianza de que, en la población, la probabilidad real estaría entre 0,5235 y 0,5931, todos números superiores al 0,50 que esperaríamos si los teóricos de entidades hubieran elegido al azar. Por el contrario, Chiu et al. descubrieron que los individuos que no eran teóricos de entidades tuvieron un nivel de predicción significativamente menor al 0,50, con un intervalo de confianza de 0,3648 a 0,4902.

RESUMEN

Al estudiar una muestra de más de un individuo, la distribución comparativa en el proceso de prueba de hipótesis es una distribución de medias de todas las muestras posibles de tamaño igual a la cantidad de casos que se están estudiando. Podemos considerar que esa distribución describe cuál sería el resultado de a) tomar una gran cantidad de muestras, cada una con la misma cantidad de unidades seleccionadas al azar de la población de individuos y, luego b) crear una distribución de las medias de esas muestras.

La distribución de medias tiene la misma media que la población de observaciones. Sin embargo, tiene una varianza menor porque las medias muestrales tienen menos probabilidad de ser extremas que las observaciones individuales. (Los extremos de cualquier muestra tienden a equiparse con los valores centrales o los valores extremos en dirección opuesta). Específicamente, la varianza de la distribución de medias es la varianza de la población de observaciones individuales dividida por la cantidad de individuos que forma cada muestra (el desvío estándar es la *raíz cuadrada de la varianza*). La forma de la distribución de medias se aproxima a la curva normal si a) la población de individuos sigue una curva normal o b) las muestras tienen 30 registros cada una, o más.

Las pruebas de hipótesis que involucran una sola muestra de más de un individuo y una población conocida se realizan de la misma forma que las pruebas de hipótesis presentadas en el capítulo 6 (donde los estudios se realizaban con un sólo individuo comparado con una población de individuos). La excepción principal es que la distribución comparativa es una distribución de medias.

La mejor estimación puntual de la media poblacional es la media muestral. Podemos determinar una estimación por intervalo de la media poblacional basándonos en la distribución de medias. Cuando la distribución de medias sigue una curva normal, el intervalo del 95% de confianza incluye todos los números, desde 1,96 desvíos estándar por debajo de la media muestral (límite de confianza inferior) hasta 1,96 desvíos estándar por encima de la media muestral (límite superior de confianza). El intervalo del 95% de confianza es un intervalo de valores acerca del cual tenemos un 95% de seguridad de que incluye la verdadera media poblacional.

Uno de los aspectos del debate actual acerca de las pruebas de significación plantea si los investigadores deberían reemplazarlas por los intervalos de confianza. Aquellos que proponen los intervalos de confianza sostienen que éstos brindan información adicional, se concentran en la estimación y reducen la utilización incorrecta propia de las pruebas de significación. Sin embargo, los intervalos de confianza rara vez se utilizan en las publicaciones científicas psicológicas, en parte, debido a la costumbre y a la falta de familiaridad con ellos, así como también a la incomodidad que presenta su descripción. Además, aquellos que se oponen a basarse exclusivamente en los intervalos de confianza sostienen que los intervalos no pueden utilizarse en algunos procedimientos avanzados, que la estimación no siempre es el objetivo deseado y que también los intervalos pueden utilizarse de formas incorrectas propias de ellos.

El tipo de prueba de hipótesis descripta en este capítulo rara vez se utiliza en la investigación práctica (la hemos aprendido como escalón hacia otros temas). El desvío estándar de la distribución de medias, con frecuencia denominado "error estándar" (*SE*), en ocasiones se utiliza para describir la variabilidad esperada de las medias, particularmente en gráficos de barra en los que el error estándar puede representarse por la longitud de un segmento ubicado sobre o debajo de la parte superior de cada barra.

Términos clave

- Intervalo de confianza (*IC*).
- Límites de confianza.
- Distribución de medias.
- Estimación por intervalos.
- Media de una distribución de medias (m*M*).
- Intervalo del 95% de confianza.
- Intervalo del 99% de confianza.
- Estimación puntual.
- Forma de la distribución de medias.
- Desvío estándar de una distribución de medias (σ_M).
- Error estándar de la media (*SE*).
- Varianza de una distribución de medias (σ^2_M).
- Prueba Z.

Ejercicios

Los ejercicios implican la realización de cálculos (con la ayuda de una calculadora). La mayoría de los problemas estadísticos reales se resuelven por computadora, pero aunque exista la posibilidad de utilizarla, es conveniente realizar estos ejercicios manualmente para incorporar el método de trabajo.

Para adquirir práctica en la utilización de una computadora, para resolver problemas estadísticos, se puede utilizar la sección de computación de cada capítulo, publicada en la *Guía de estudio y libro de tareas de computación para el alumno [Student's Study Guide and Computer Workbook]* que acompaña este libro.

Todos los datos de esta sección son ficticios (a menos que se especifique lo contrario).

Las respuestas a los ejercicios de la serie I se encuentran al final del libro.

SERIE I

1. Explique por qué el desvío estándar de la distribución de medias generalmente es menor que el desvío estándar de la distribución poblacional de observaciones individuales.

2. En el caso de una población de observaciones individuales que tiene un desvío estándar de 10, ¿cuál es el desvío estándar de la distribución de medias de muestras de los siguientes tamaños: a) 2, b) 3, c) 4, d) 5, e) 10, f) 20 y g) 100?

3. Para cada uno de los casos del problema 2 calcule el intervalo del 95% de confianza (es decir, los límites de confianza superior e inferior). Suponga en cada caso que la muestra del investigador tiene una media de 100 y que la población sigue una distribución normal.

4. Cierta población de individuos tiene una media de 40, un desvío estándar de 6, y sigue una distribución normal. Indique si cada una de las siguientes muestras está entre las que tendrían menos del 5% de probabilidad de ser seleccionadas al azar de esa población: a) una muestra de 10 con una media de 44; b) una muestra de 1 con una media de 48; c) una muestra de 81 con una media de 42; y d) una muestra de 16 con una media de 42. En cada caso

a) muestre los cálculos por los cuales llegó a su respuesta y b) incluya un diagrama de las distribuciones involucradas.

5. Veinticinco mujeres de entre 70 y 80 años de edad fueron seleccionadas al azar de la población general de mujeres de esa edad para participar en un programa especial para disminuir el tiempo de reacción. Después del curso, las mujeres tenían un tiempo de reacción promedio de 1,5 segundos. Suponiendo que el tiempo de reacción medio para la población general de mujeres de esa edad es 1,8, con un desvío estándar de 5 segundos (y que además la población es aproximadamente normal), ¿cuál sería su conclusión acerca de la eficacia del curso? a) Siga los pasos de la prueba de hipótesis (utilice el nivel 0,01). b) Calcule el intervalo del 99% de confianza. c) Explique su respuesta a alguien que está familiarizado con la lógica general de la prueba de hipótesis, la curva normal, las puntuaciones Z y la probabilidad, pero que no está familiarizado con la idea de una distribución de medias o de un intervalo de confianza.

6. Una gran cantidad de personas observó un filme sobre un accidente automovilístico entre un automóvil en movimiento y un automóvil detenido. Cada persona llenó luego un cuestionario sobre cuán verosímil es que el conductor del automóvil en movimiento tuviera la culpa, conforme a una escala que iba desde **no tuvo la culpa** = 0 hasta **fue completamente culpable** = 10. La distribución de las puntuaciones en condiciones ordinarias sigue una distribución normal, $\mu = 5{,}5$, y $\sigma = 0{,}8$. Se analizan las respuestas de dieciséis individuos seleccionados al azar, a quienes se les cambió la redacción de la pregunta. En estas distintas condiciones, la pregunta es: ¿Cuán verosímil es que el conductor del auto que se estrelló contra el otro fuera el culpable? (La diferencia radica en que, en estas condiciones, en lugar de describir el hecho en forma neutra, la pregunta utiliza la frase "se estrelló"). Utilizando estas instrucciones diferentes, los 16 participantes dieron una media, en cuanto a la puntuación de culpabilidad, de 5,9. ¿El cambio de instrucciones aumentó significativamente la puntuación media de culpabilidad? a) Siga los pasos de la prueba de hipótesis (utilice el nivel 0,05). b) Calcule el intervalo del 95% de confianza. c) Explique su respuesta a alguien que nunca ha estudiado estadística.

7. Corte 90 papeles pequeños y escriba 10 veces los números del 1 al 9, una en cada papel. Ponga los papeles en un recipiente grande y mézclelos. Ahora saque un papel, escriba el número que se lee en el papel y regréselo al recipiente. Realice este mismo procedimiento 20 veces. Cree un histograma y calcule la media y la varianza del resultado. Debería obtener una distribución aproximadamente rectangular. Después tome dos papeles, calcule su media, anótela y vuelva a colocar los papeles en el recipiente.[6] Repita este proceso unas 20 veces. Cree un histograma y luego calcule la media y la varianza de esta distribución de medias. La varianza debería ser aproximadamente la mitad de la varianza anteriormente calculada. Finalmente, repita el proceso nuevamente, pero esta vez sacando tres papeles por vez. La distribución de medias de tres papeles por vez debería tener una varianza de aproximadamente un tercio de la distribución de muestras de un papel cada una. Observe también que, a medida que aumenta el tamaño de la muestra, sus distribuciones se acercan a lo normal. (Si hubiera comenzado con una distribución de papeles *normalmente* distribuida, las distribuciones de medias hubieran estado bastante cerca de lo normal, independientemente de la cantidad de papeles de cada muestra).

[6] Técnicamente, al sacar las muestras de dos papeles, debería hacerlo sacando una, anotando el número y poniéndola luego con las demás. Luego, sacando la otra, escribiendo el número y poniéndola nuevamente con las demás. Estas dos observaciones se considerarían una muestra de la cual se calcularía la media. Lo mismo se aplica a muestras de tres papeles. El proceso descripto se denomina muestreo con **reemplazo**. Sin embargo, con 90 papeles en un recipiente, sacar dos o tres papeles al mismo tiempo y ponerlos nuevamente con los demás será una aproximación bastante cercana para este ejercicio y le ahorrará algo de tiempo.

SERIE II

1. ¿En qué condiciones es razonable suponer que una distribución de medias seguirá una distribución normal?

2. Indique la media y el desvío estándar de la distribución de medias de cada una de las siguientes situaciones:

	Población		Tamaño Muestra
	Media	*Varianza*	
(a)	100	40	10
(b)	100	30	10
(c)	100	20	10
(d)	100	10	10
(e)	50	10	10
(f)	100	40	20
(g)	100	10	20

3. Para cada uno de los ejemplos anteriores, calcule el intervalo del 95% de confianza, suponiendo que el investigador tenía una media muestral de 80 en cada caso (y que las poblaciones siguen una distribución normal).

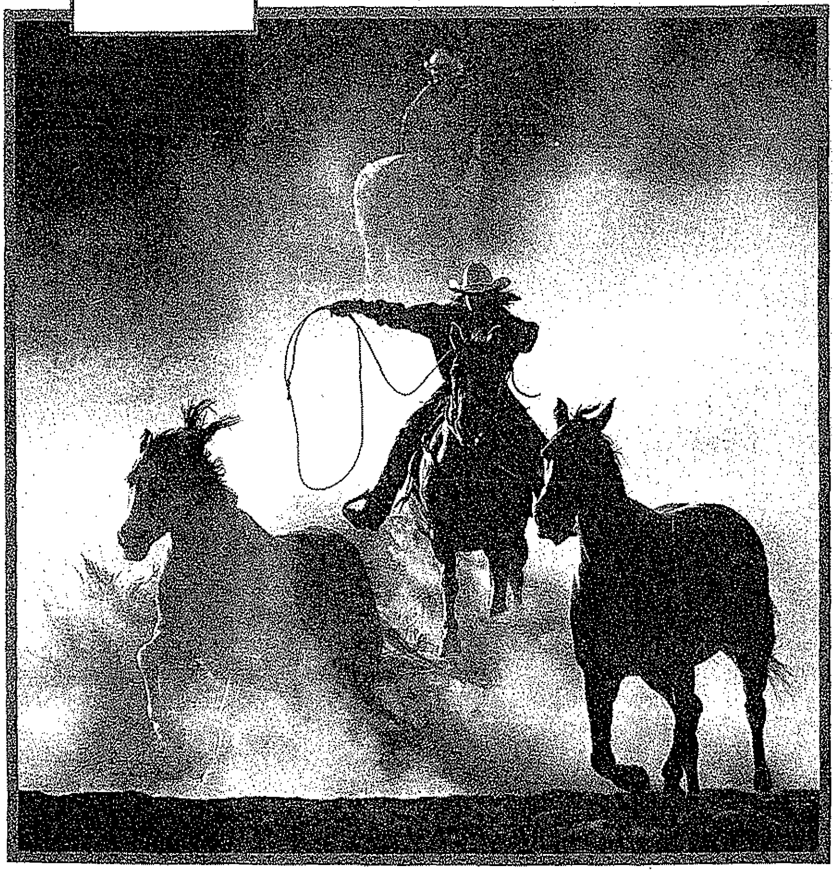
4. Basándose en la información dada, establezca su conclusión para cada estudio. (Asegúrese de indicar las características de la distribución comparativa, el punto de corte, el valor de la media muestral en la distribución comparativa y su conclusión en cuanto al rechazo o no de la hipótesis nula. Todas las pruebas de hipótesis son de dos colas).

	Población		Tamaño muestral	Media muestral	Nivel de significación
	μ	σ			
(a)	36	8	16	38	0,05
(b)	36	6	16	38	0,05
(c)	36	4	16	38	0,05
(d)	36	4	16	38	0,01
(e)	34	4	16	38	0,01

5. Un investigador está interesado en averiguar si las personas son capaces de identificar correctamente las emociones de personas de otras culturas. Se sabe que utilizando determinado método de medición, las posiciones en cuanto a la precisión de los adultos norteamericanos en general están distribuidas normalmente con una media de 82 (de un total de 100) y una varianza de 20. La distribución se basa en las posiciones obtenidas al identificar las emociones expresadas por miembros de su propia cultura (otros norteamericanos). En el estudio que estamos analizando, el investigador organiza a 50 adultos norteamericanos para que identifiquen las emociones de individuos de Indonesia. La precisión media de estos 50 individuos fue 78. Utilizando un nivel de 0,05, ¿cuál debería ser la conclusión del investigador? a) Siga los pasos de la prueba de hipótesis. b) Calcule el intervalo del 95% de confianza. c) Explique su respuesta a alguien que nunca ha estudiado estadística.

6. Un psicólogo está interesado en las condiciones que afectan la cantidad de sueños que las personas recuerdan por mes y en los cuales se encuentran solos. Supondremos que, basándonos en previas investigaciones extensivas, se sabe que en la población general la cantidad de tales sueños por mes sigue una distribución normal, con $\mu = 5$ y $\sigma = 4$. El investigador desea probar la predicción que establece que la cantidad de sueños como los descriptos será mayor entre aquellas personas que recientemente hayan experimentado un hecho traumático. Por lo tanto, el psicólogo analiza 36 individuos que han experimentado recientemente un hecho traumático, haciéndoles llevar un registro de sus sueños durante un mes. La media de sueños en los que se encuentran solos es 8. ¿Llegaría usted a la conclusión de que las personas que han sufrido recientemente una experiencia traumática tienen una cantidad significativamente diferente de sueños en los que se encuentran solas? a) Siga los pasos de la prueba de hipótesis (utilice el nivel 0,05). b) Calcule el intervalo del 95% de confianza. c) Explique su respuesta a alguien que está familiarizado con la lógica general de la prueba de hipótesis, la curva normal, las puntuaciones Z y la probabilidad, pero que no está familiarizado con la idea de una distribución de medias o intervalos de confianza.

8 Potencia estadística y tamaño del efecto



Descripción del capítulo

- ¿Qué es la potencia estadística?
- Alfa, beta y potencia.
- Cálculo de la potencia estadística.
- Tablas de potencia.
- ¿Qué factores determinan la potencia de un estudio?
- Tamaño del efecto.
- Tamaño de la muestra.
- Otros factores que influyen en la potencia.
- Papel que desempeña la potencia al diseñar un experimento.
- La importancia de la potencia en la evaluación de los resultados de un estudio.
- Potencia, tamaño del efecto e intervalos de confianza.
- Meta-análisis.
- Controversias y limitaciones: continuación de la controversia acerca de la significación estadística: tamaño del efecto versus significación estadística.
- Potencia y tamaño del efecto según se describen en publicaciones científicas.
- Resumen.
- Términos clave.
- Ejercicios.

Potencia es la capacidad para cumplir objetivos. Por eso, una medida razonable de potencia en cualquier situación dada es la probabilidad de cumplir con los objetivos en esa determinada situación. El objetivo de un investigador que realiza un experimento es la obtención de un resultado significativo, siempre que la hipótesis de investigación realmente sea verdadera. La **potencia estadística** de un estudio es la probabilidad de que ese estudio tenga un resultado significativo si la hipótesis de investigación es verdadera.

Calcular la potencia al planificar un estudio ayuda a definir la cantidad de participantes que se van a utilizar. Además, comprender el concepto de potencia es sumamente importante para cualquiera que lea publicaciones de investigación psicológica; por ejemplo, para comprender los resultados experimentales que no son significativos o resultados que son significativos estadísticamente pero no en la práctica.

En este capítulo, examinamos sistemáticamente el concepto de potencia estadística. Qué es, cómo se calcula, qué factores influyen en ella, y por qué es importante. Es nuestra obligación advertir que, a veces, este material acerca de la potencia puede resultar particularmente difícil de captar. Pero vale la pena aprenderlo. Por eso, recomendamos al lector ser paciente consigo mismo y tomarse todo el tiempo que sea necesario. Estamos seguros de que lo logrará.

Como parte del proceso de aprendizaje de la potencia, el capítulo también presenta la noción de tamaño del efecto. Como veremos, el tamaño del efecto es un punto crucial para comprender la potencia, y un tema de considerable importancia en sí mismo para comprender las investigaciones psicológicas.

¿QUÉ ES LA POTENCIA ESTADÍSTICA?

Dijimos que la potencia estadística de un experimento es la probabilidad de que el estudio arroje un resultado significativo si la hipótesis de investigación es verdadera. Es importante tener en cuenta que la potencia de un experimento implica determinada situación, si la hipótesis de investigación es verdadera. No nos interesa lograr un resultado significativo si la hipótesis de investigación es falsa.

Ahora bien, podríamos preguntarnos lo siguiente: "¿Si la hipótesis de investigación es verdadera, no dará el experimento automáticamente un resultado significativo?" La respuesta es no; puede ocurrir que la muestra particular que fue seleccionada de la población no resulte lo suficientemente extrema como para rechazar la hipótesis nula.

Ejemplo

Analicemos nuevamente el ejemplo del capítulo 7 acerca de las instrucciones especiales a alumnos de quinto grado que están dando un examen estándar de nivel. En el proceso de prueba de hipótesis de este ejemplo comparamos dos poblaciones:

Población 1: alumnos de quinto grado que reciben instrucciones especiales.
Población 2: alumnos de quinto grado que no reciben instrucciones especiales.

La hipótesis de investigación establecía que la población 1 tendría puntuaciones más altas que la población 2. $H_0: \mu_1 > \mu_2$

La distribución superior de la figura 8-1 grafica la situación en la que la hipótesis de investigación es verdadera. La distribución inferior representa a la población 2. Dado que estamos interesados en medias de muestras formadas por 64 individuos, ambas distribuciones son distribuciones de medias.

La distribución inferior es también la distribución comparativa, es decir, la distribución de medias que esperaríamos para ambas poblaciones si la hipótesis nula fuera verdadera. El área sombreada en la cola derecha de la distribución inferior es el área en la cual rechazaríamos la hipótesis nula si, como resultado del estudio, la media muestral se encontrara bajo esa área. El área de rechazo sombreada comienza a 209,84 (una puntuación Z de 1,64) y abarca un 5% de la distribución comparativa.

La distribución de medias superior es la que **predicen** los investigadores para la población que recibe instrucciones especiales (población 1). En el capítulo 7, nunca hablamos de esa distribución, en parte porque la distribución de la población predicha es bastante imaginaria, **a menos que** la hipótesis de investigación sea verdadera. Si la hipótesis nula es verdadera, la distribución de la población 1 sería igual a la distribución que se basa en la población 2. Es decir, si la hipótesis nula es verdadera, la distribución de la población 1 no estaría desplazada hacia la derecha.

No obstante, para aprender el tema de la potencia, aquí analizamos la situación en la cual la hipótesis de investigación es verdadera. En esa situación, la media de la población 1 se encuentra más hacia la derecha que la media de la población 2 (distribución comparativa). Es decir, en esa situación, las puntuaciones en el examen de nivel son, en promedio, mayores en la población 1 que en la población 2. Específicamente, la distribución de medias superior (población 1 predicha) tiene una media de 208; la media de la distribución comparativa es sólo de 200, lo que muestra que se espera que la población que recibe las instrucciones especiales (población 1) tenga una media 8 puntos mayor.

Supongamos ahora que los psicólogos expertos en educación realizan el experimento. Éstos dan las instrucciones especiales a un grupo de 64 alumnos de quinto grado y calculan la puntuación media en el examen. Supongamos que la hipótesis de investigación es verdadera. Recordemos que *si la hipótesis de investigación es verdadera*, la media del grupo de 64 alumnos de quinto grado pertenece a una distribución semejante a la curva superior en la figura 8-1.

En este ejemplo, sin embargo, la distribución superior de medias (tomada de la predicción del investigador sobre la población 1) se encuentra sólo levemente volcada hacia la derecha de la distribución comparativa. Es decir, los psicólogos predicen sólo un pequeño aumento de los registros (ocho puntos) a causa de las instrucciones especiales; por lo tanto, la distribución superior se encuentra desplazada sólo una pequeña distancia hacia la derecha en comparación con la distribución inferior, que es la distribución comparativa. Lo que la figura nos indica es que cualquier

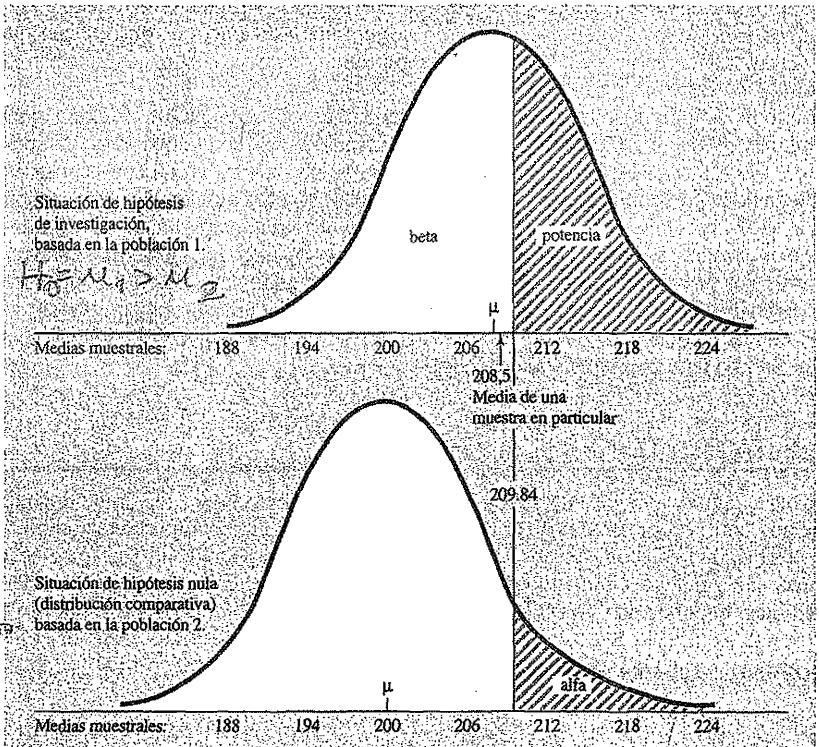


Figura 8-1. Distribuciones de las puntuaciones medias de exámenes rendidos por 64 alumnos de quinto grado, tomadas de un estudio ficticio de alumnos de quinto grado rindiendo un examen estándar de nivel. La distribución de medias inferior se basa en una distribución conocida de puntuaciones individuales de alumnos de quinto grado que no recibieron ninguna instrucción especial en cuanto al examen (población 2). La distribución de medias superior se basa en una distribución predicha de puntuaciones individuales de alumnos de quinto grado que recibieron instrucciones especiales en cuanto al examen (población 1). Los investigadores predicen una media de 208 para esta población. Las áreas sombreadas de ambas distribuciones muestran el área bajo la cual será rechazada la hipótesis nula.

media tomada de la distribución superior no estará, probablemente, lo *suficientemente* volcada hacia la derecha en la distribución inferior como para rechazar la hipótesis nula. De hecho, menos de la mitad de la distribución superior está sombreada. Para decirlo de otro modo, si la hipótesis de investigación es verdadera, la muestra que estudiamos es, en efecto, una muestra aleatoria de la distribución que aquí aparece como población 1. Sin embargo, las chances de que una muestra aleatoria de esa población se encuentre bajo el área sombreada están por debajo del 50%.

Por ejemplo, supongamos que la muestra analizada, formada por 64 alumnos de quinto grado, tenía una media de 208,5, tal como lo indica la flecha en la figura. Dado que es necesario una media de al menos 209,84 para rechazar la hipótesis nula, el resultado de este experimento no sería significativo. No lo sería, aun cuando la hipótesis de investigación en realidad es verdadera (y aun cuando el valor muestral es mayor que la media de la distribución comparativa).

Indudablemente, es posible que los investigadores pudieran seleccionar una muestra de la población 1 con una media lo suficientemente alejada hacia la derecha como para ubicarse bajo el área sombreada (es decir, con un promedio en el examen lo suficientemente alto). Sin embargo, dada la forma en la que hemos establecido el ejemplo, las chances de que el experimento no resulte significativo son más del 50%, aun cuando sabemos que la hipótesis de investigación es verdadera.

Cuando un estudio tiene chances de menos del cincuenta por ciento de resultar significativo, aun si la hipótesis de investigación es verdadera, decimos que el estudio tiene baja potencia. Pero, ¿qué sucedería si la situación fuera tal que se esperara que la curva superior estuviera corrida bien a la derecha de la curva inferior, es decir, que la predicción fuera que aquellos que rinden el examen con las instrucciones especiales tendrán puntuaciones realmente altas? (La figura 8-3 que aparece más adelante en el capítulo es un ejemplo de este tipo de situación). En la situación descripta, la mayor parte bajo la curva superior estaría sombreada, y casi cualquier muestra tomada de esa curva superior estaría bajo el área de rechazo de la curva inferior. Lo anterior significa que cuando uno realiza *el estudio, la probabilidad de obtener un resultado significativo sería alta, por* lo cual el estudio tendría potencia alta.

ALFA, BETA Y POTENCIA

Al analizar la significación estadística y la potencia, es útil pensar en función de tipos de "errores" que uno podría cometer al utilizar la prueba de significación. Es importante comprender que no estamos hablando de cometer errores con los cálculos ni tampoco de utilizar los procedimientos equivocados. Estamos hablando de que, incluso cuando hacemos todo adecuadamente, aun podemos sacar conclusiones erróneas. Es decir, estamos hablando de error en el sentido de obtener un resultado equivocado de un procedimiento correcto. En ese sentido, podemos cometer dos tipos de errores: error Tipo I y error Tipo II.[1]

Error Tipo I y alfa

Supongamos que realizamos un estudio y establecemos el corte del nivel de significación en un nivel de probabilidad muy alto, como por ejemplo del 20%. En esas condiciones, la hipótesis nula sería rechazada muy fácilmente. Si realizáramos muchos estudios de este tipo, con frecuencia (aproximadamente un 20% de las veces) decidiríamos que la hipótesis de investigación se susten-

[1] Ocasionalmente puedes llegar a escuchar mencionar el error Tipo III. Se trata de llegar a la conclusión de que existe un resultado significativo en una determinada dirección cuando el efecto real es en la dirección opuesta.

ta cuando, en realidad, no deberíamos hacerlo. A esto se lo denomina **error Tipo I**. La probabilidad de cometer un error Tipo I, que se denomina **alfa**, es el nivel de significación. Por lo tanto, en la mayoría de los estudios, alfa es igual a 0,05.

Al realizar investigaciones, nunca sabemos a ciencia cierta si la hipótesis de investigación o la hipótesis nula son verdaderas. Los resultados del procedimiento de prueba de hipótesis pueden o no llevarnos a rechazar la hipótesis nula, pero, en cualquier caso, no estamos seguros de haber tomado la decisión correcta.

Supongamos que las instrucciones especiales, en nuestro ejemplo de los alumnos de quinto grado, en realidad no produjeron ninguna diferencia, y que la hipótesis nula era verdadera. Supongamos además que al realizar el estudio, simplemente sucedió que los investigadores seleccionaron, para recibir las nuevas instrucciones, a algunos alumnos que eran inusualmente buenos en ese tipo de examen. Aunque es poco probable, podría suceder, y el efecto sería que los investigadores rechazarían la hipótesis nula y concluirían que las instrucciones especiales producen una diferencia. Esta decisión de rechazar la hipótesis nula sería equivocada, error Tipo I. Por supuesto, los investigadores no podrían saber que cometieron un error de este tipo. La seguridad que tienen los investigadores es saber que la probabilidad de cometer tal error es baja (menos del 5% si utilizamos el nivel de significación de 0,05).

Los errores Tipo I son una gran preocupación para los investigadores psicológicos, quienes podrían construir teorías completas y programas de investigación, para no mencionar aplicaciones prácticas, sobre la base de conclusiones derivadas de pruebas de hipótesis que en realidad están equivocadas. Debido a que estos errores son tan preocupantes, se los denomina Tipo I.

Como ya hemos señalado, los investigadores no pueden saber cuándo han cometido un error Tipo I; no obstante, pueden intentar realizar estudios en los que las posibilidades de cometer un error Tipo I sean lo más pequeñas posibles. Supongamos que para determinado estudio establecemos el nivel de significación en $p<0,05$, que indica que rechazaremos la hipótesis nula si existe menos de un 5% (0,05) de probabilidad de que pudiéramos haber obtenido nuestro resultado si la hipótesis nula fuera verdadera. Al rechazar la hipótesis nula en esas circunstancias, estamos admitiendo hasta un 5% de probabilidades de obtener nuestro resultado aun cuando la hipótesis nula fuera realmente verdadera. Es decir, estamos admitiendo un 5% de probabilidad de cometer un error Tipo I: alfa es igual al 5%.

Podríamos disminuir alfa haciendo aún menos probable el rechazo de la hipótesis nula por error. Por ejemplo, utilizar un nivel 0,001 de significación sería como contratar un seguro contra el error Tipo I. En ese caso, habría menos de una posibilidad en mil de cometer el error Tipo I. Sin embargo, al igual que cuando contratamos un seguro, a mayor protección, más alto es el costo. Existe un costo que pagar por establecer un nivel de significación a un nivel demasiado extremo. A continuación hablaremos acerca de ese costo.

Error tipo II y beta

Si establecemos un nivel de significación muy riguroso, como por ejemplo 0,001, corremos otro tipo de riesgo. En ese caso, podemos realizar un estudio en el que la hipótesis de investigación es verdadera pero el resultado no es lo suficientemente extremo como para rechazar la hipótesis nula. Por lo tanto, el error que cometeríamos sería no rechazar la hipótesis nula cuando, en realidad, la hipótesis nula es falsa. Este es el **error Tipo II**. La probabilidad de cometer un error Tipo II se denomina **beta**. (No debemos confundir esta beta con el coeficiente de regresión estandarizado que explicamos en el capítulo 4, al que también llamamos beta).

Analicemos nuevamente el ejemplo sobre los alumnos de quinto grado. Supongamos que, en verdad, dar las instrucciones especiales **sí** hace que los alumnos de quinto grado se desempeñen mejor en el examen. Sin embargo, al realizar el estudio, los resultados no muestran ese patrón. Tal vez, la muestra que seleccionamos al azar para probar las nuevas instrucciones incluían principalmente alumnos de quinto grado con un nivel particularmente bajo para este tipo de examen. Como hemos visto, aun cuando el procedimiento pueda haberlos ayudado a lograr un mejor desempeño, de todos modos sus puntuaciones pueden no ser mucho más altas que el promedio de los alumnos de quinto grado que no recibieron instrucciones especiales. Los resultados no serían significativos. Por lo tanto, haber decidido no rechazar la hipótesis nula, y rehusarse a sacar una conclusión, sería un error Tipo II. Los errores Tipo II preocupan especialmente a los psicólogos interesados en aplicaciones prácticas, ya que un error de este tipo podría provocar que no se implemente un procedimiento práctico útil.

Al igual que con los errores Tipo I, no podemos saber cuándo hemos cometido un error Tipo II. Pero podemos intentar realizar nuestros estudios de forma tal de reducir la probabilidad de cometer un error Tipo II. Una forma de contratar un seguro contra un error Tipo II es establecer un nivel de significación muy indulgente, como por ejemplo $p<0{,}10$ ó incluso $p<0{,}20$. De ese modo, aun cuando un estudio arroje una diferencia muy pequeña, hay muchas chances de que los resultados sean significativos. No obstante, también hay que pagar un costo por esta póliza de seguros. El costo es correr demasiado riesgo de cometer un error Tipo I.

Un error Tipo II ocurre *cuando decidimos que el experimento no es concluyente* (no rechazamos la hipótesis nula) y en realidad nuestra hipótesis de investigación era verdadera. En ese caso, el experimento no sustentó la hipótesis de investigación cuando debería haberlo hecho. Esta es la situación ilustrada en la figura 8-1. En ese ejemplo, se cometió un error Tipo II.

En la figura 8-1, beta es el área no sombreada de la distribución superior (distribución que se basa en la predicción de la hipótesis de investigación con respecto a la población 1). Es el área donde, aun cuando la hipótesis de investigación sea verdadera, una media no sería lo suficientemente extrema como para que podamos rechazar la hipótesis nula; es el área de la distribución superior que se encuentra a la izquierda del punto en el que comienza el área alfa en la distribución inferior (comparativa).

La **potencia** de un experimento es la probabilidad de que si la hipótesis de investigación es verdadera, el experimento la sustente (rechace la hipótesis nula). Es decir, potencia es la probabilidad de **no** cometer un error Tipo II. Numéricamente, potencia es 1 menos beta. En la figura 8-1, la potencia es la porción sombreada de la distribución superior. En este ejemplo (el 50% del área bajo la curva), es menor a 0,5.

Relación entre los errores Tipo I y Tipo II

Al momento de establecer los niveles de significación, protegerse contra un tipo de error aumenta las chances de cometer el otro tipo de error. El costo de la póliza de seguros contra el error Tipo I (establecer un nivel de significación de, digamos, 0,001) es aumentar beta, la probabilidad de cometer el error Tipo II. (Esto ocurre porque con un nivel de significación extremo como 0,001, aun si la hipótesis de investigación es verdadera, los resultados deben ser demasiado contundentes para ser lo suficientemente importantes como para rechazar la hipótesis nula). El costo de la póliza de seguros contra el error Tipo II (establecer un nivel de significación de, digamos, 0,20) es aumentar las posibilidades de cometer el error Tipo I. (Esto ocurre porque con un nivel de significación como 0,20, aun si la hipótesis nula fuera verdadera, es bastante fácil obtener un resultado significativo sólo por haber seleccionado accidentalmente una muestra que, aun antes de realizar el estudio, tenía un nivel mayor o menor que la población general).

La negociación entre estos dos temas conflictivos se resuelve usualmente por convención; a eso se deben los niveles de significación estándar del 5% y el 1%.

Tabla 8-1.
Posibles decisiones correctas y erróneas en la prueba de hipótesis.

		Condición real de la hipótesis de investigación (en la práctica, desconocido)	
		Verdadera	*Falsa*
Decisión al aplicar el procedimiento de prueba de hipótesis	*Se sostiene la hipótesis de investigación (se rechaza la hipótesis nula)*	Decisión correcta; p = potencia	Error tipo I; p = alfa
	El estudio no es concluyente (no se puede rechazar la hipótesis nula)	Error tipo II; p = beta	Decisión correcta; $p = 1 -$ alfa

Visión general de los posibles resultados de la prueba de hipótesis teniendo en cuenta alfa, beta y la potencia

La tabla 8-1 diagrama las posibles decisiones correctas y erróneas en la prueba de hipótesis. En la parte superior de la tabla encontramos las dos posibilidades en cuanto a la veracidad o no de la hipótesis de investigación. (Esto nunca lo sabemos realmente). En el costado se plantea si, después de realizar la prueba de hipótesis, decidimos que los resultados a) sostienen la hipótesis de investigación (rechazan la hipótesis nula) o b) no son concluyentes (no rechazan la hipótesis nula).

Luego, la tabla muestra los dos modos de tomar la decisión correcta y los dos modos de cometer el error que hemos analizado en esta sección. También muestra los nombres de la probabilidad de cada una de las cuatro posibles decisiones correctas o erróneas.

CÁLCULO DE LA POTENCIA ESTADÍSTICA

La potencia de un experimento se puede calcular. En el ejemplo de los alumnos de quinto grado, calcular la potencia implica calcular el área de la porción sombreada en la distribución superior de la figura 8-1. Las distribuciones de medias son, por lo general, aproximadamente normales, y este es claramente el caso del ejemplo que estamos analizando (ya que el tamaño de la muestra es mayor a 30). Por lo tanto, para realizar los cálculos de la potencia utilizaremos puntuaciones Z. Primero, calculamos la puntuación Z correspondiente al punto en el que comienza el área sombreada en la distribución superior; luego, determinamos el área correspondiente a través de la tabla de áreas bajo la curva normal.

Ejemplo

Volvamos a analizar la figura 8-1 que representa gráficamente las distribuciones de medias del ejemplo de los alumnos de quinto grado. La población de individuos que no recibe instrucciones especiales tenía una media de 200 y un desvío estándar de 48 (una varianza de 2.304). Los investigadores analizaron una muestra de 64 alumnos de quinto grado. De ese modo, en el capítulo 7 calculamos que el desvío estándar de la distribución de medias es 6 (es decir, $\sqrt{2.304/64} = 6$). Anteriormente, en este mismo capítulo, dijimos que los investigadores predijeron que las instrucciones especiales aumentarían la media a 208. La figura 8-2 muestra las puntuaciones Z correspondientes a ambas distribuciones, basándose en estos números.[2]

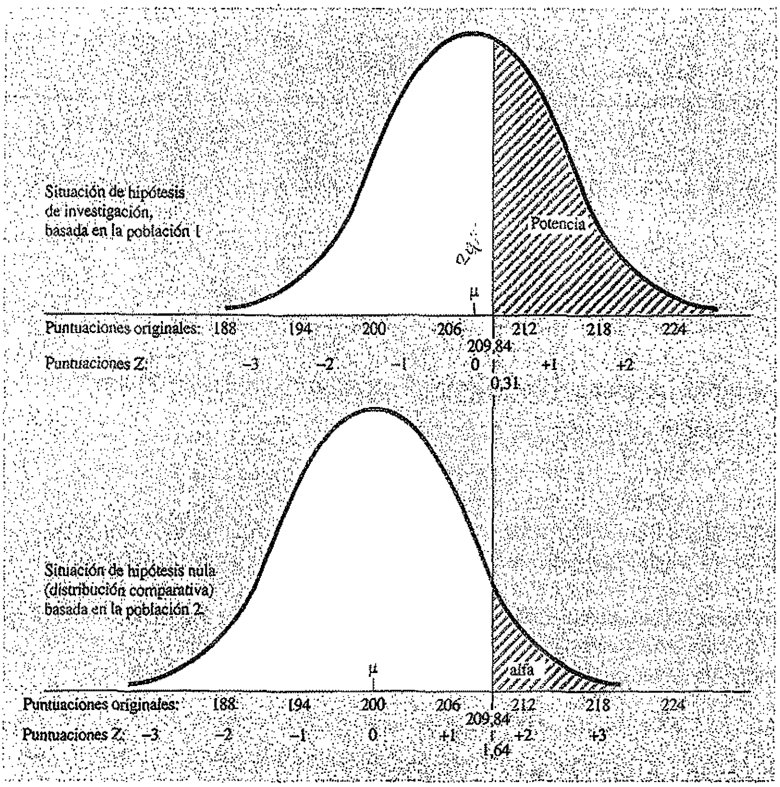


Figura 8-2. Distribuciones de medias de 64 resultados de exámenes, basadas en distribuciones predichas (superior) y conocidas (inferior) de un estudio ficticio de alumnos de quinto grado que reciben instrucciones especiales antes de rendir un examen estándar de nivel. En ambas distribuciones se indican las puntuaciones Z y las puntuaciones originales correspondientes al punto de corte basado en la distribución inferior. (El punto de corte corresponde a un nivel de significación de $p<0{,}05$, prueba de una cola).

[2] Normalmente suponemos que, independientemente de que la hipótesis nula sea verdadera (es decir, si las medias de las dos poblaciones son iguales), las varianzas de ambas poblaciones serán iguales. Las distribuciones de medias de ambas poblaciones también se basan en la misma cantidad de observaciones en cada muestra (en este ejemplo 64). Por lo tanto, los desvíos estándar de estas dos distribuciones de medias también serán iguales.

En el capítulo 7 determinamos que, utilizando un nivel de significación del 5%, con una prueba de una cola, para rechazar la hipótesis nula necesitamos que la puntuación Z correspondiente a la media muestral sea de, al menos, 1,64. Utilizando la fórmula para convertir puntuaciones Z en puntuaciones originales, la puntuación Z determinada corresponde a una puntuación original de 209,84, es decir, 200 + (1,64 x 6) = 209,84.

Como ya dijimos, los investigadores predijeron una media de 208 para los alumnos de quinto grado que reciben instrucciones especiales (población 1). El punto de corte de 209,84 está 1,84 puntos de prueba por encima de la media general de 208 de esa distribución, dando una puntuación Z de 0,31 (es decir, 1,84/6=0,31).

La tabla de áreas bajo la curva normal muestra que un 12% del área se encuentra entre la media y una Z de 0,31. Por lo tanto, un 38% supera a la Z de 0,31. En otras palabras, un 38% de la distribución de medias predicha para la población 1 se encuentra por encima de una puntuación Z de 0,31 (y por lo tanto el 38% de las medias se encuentran por encima de la puntuación original 209,84).

La conclusión es la siguiente: suponiendo que la predicción de los investigadores sea correcta, tienen sólo un 38% de chances de que la muestra de 64 alumnos que analizaron arroje una media lo suficientemente alta como para que el resultado sea significativo. Es decir, existe sólo un 38% de chances de obtener una media mayor a 209,84, aun suponiendo que la hipótesis de investigación sea verdadera. Por lo tanto, decimos que la potencia de este experimento es del 38%. Beta, la probabilidad de cometer un error Tipo II, es del 62% (es decir, 100% – 38% = 62%).

Es importante observar que la forma en la que calculamos la potencia no tiene nada que ver con el resultado real del estudio. De hecho, los investigadores por lo general calculan la potencia antes de realizar el estudio.

Resumen de los pasos para el cálculo de la potencia

En las condiciones del ejemplo que estamos analizando (la media de una sola muestra comparada con una población conocida), calcular la potencia incluye cuatro pasos:

1. Reunir la información necesaria: a) la media y el desvío estándar de la población 2 (distribución comparativa) y b) la media predicha de la población 1 (población que recibió el procedimiento experimental). También resultará muy útil crear un diagrama de las dos distribuciones de modo similar a la figura 8-2.
2. Determinar, en la distribución comparativa, el punto de corte para rechazar la hipótesis nula.
3. Determinar la puntuación Z del punto de corte anterior, pero en la distribución de medias de la población que recibe la manipulación experimental.
4. Utilizando la tabla de áreas bajo la curva normal, determinar la probabilidad de obtener un registro más extremo que esa puntuación Z.[3]

Otro Ejemplo

Analicemos otro ejemplo ficticio. Una gran empresa está intentando decidir si adopta una nueva política de promoción sanitaria. Conforme a esta nueva política, se evalúa a los empleados individualmente y se les brinda la capacitación y el asesoramiento necesarios con respecto a distintos

[3] El método descripto de cálculo de la potencia (que es el único método de cálculo de la potencia tratado en este libro) supone que las distribuciones de medias están normalmente distribuidas.

comportamientos relacionados con la salud (ejercicio, dieta, cigarrillo, etc.). Para probar la efectividad de la política, los psicólogos de la empresa planifican el siguiente estudio: se seleccionarán ochenta empleados al azar para participar del mismo, y al finalizar el año se medirá su estado general de salud conforme a una prueba estándar. La misma empresa ha realizado pruebas extensivas a sus empleados, por lo que los investigadores saben que en toda la empresa (la población de este estudio) la media en las pruebas estándar de salud es 58, el desvío estándar es 14, y los valores se distribuyen normalmente. Para que se justifique la realización del programa, debe producirse una mejora de al menos 5 puntos (es decir, la media predicha es 63). Los psicólogos de la empresa planifican utilizar un nivel de significación de 0,05.

La figura 8-3 representa gráficamente las distribuciones de medias correspondientes a las dos poblaciones involucradas en este estudio. ¿Cuál es la potencia de este experimento?

1. Reunir la información necesaria. En este ejemplo, la media de la distribución comparativa es 50. La media predicha de la población que recibe el procedimiento experimental es 63. La varianza de la población es 196 (es decir, $14^2 = 196$), por lo tanto, la varianza de la distribución de medias (distribución comparativa) es 2,45 (196/80 = 2,45), lo que nos da un desvío estándar de 1,57 ($\sqrt{2,45} = 1,57$).

2. Determinar, en la distribución comparativa, el punto de corte para rechazar la hipótesis nula. Con un nivel de significación del 5%, en una prueba de una cola, la puntuación Z de corte es +1,64. Una puntuación Z de +1,64 es igual a una puntuación original de 60,57 (es decir, 58 + [1,64 × 1,57] = 60,57). Por lo tanto, en la curva inferior (distribución comparativa) de la figura 8-3, hemos sombreado el área a la derecha del punto 60,57. Es la región alfa.

3. Determinar la puntuación Z del punto de corte anterior, pero en la distribución de medias correspondiente a la población que recibe la manipulación experimental. En esa distribución (basándonos en los valores predichos para la población 1), una puntuación original de 60,57 es igual a una puntuación Z de −1,55 (es decir, [60,57 − 63]/1,57 = −1,55). Por lo tanto, en la curva superior de la figura 8-3, hemos sombreado el área a la derecha del punto −1,55. Esa área sombreada indica la potencia del estudio, es el área sobre la zona en la que la media de una muestra real sería significativa con respecto a la distribución comparativa.

4. Utilizando la tabla de áreas bajo la curva normal, determinar la probabilidad de obtener un valor más extremo que esa puntuación Z. La tabla de áreas bajo la curva normal indica aproximadamente un 44% entre la media y una Z de 1,55. Estamos interesados en toda el área a la derecha de −1,55, por lo tanto, existe un total del 44% entre −1,55 y la media, más el 50% por encima de la media, lo que da un total de 94%. La potencia de este experimento es del 94% (beta es del 6%).

TABLAS DE POTENCIA

Los procedimientos que hemos descripto para el cálculo de la potencia se aplican cuando estamos frente a una población conocida y frente a una sola muestra. En situaciones de investigación más complejas (que analizaremos en varios de los capítulos siguientes), calcular la potencia es bastante más trabajoso. Por eso, generalmente los investigadores buscan la potencia de un estudio utilizando cuadros especiales, denominados **tablas de potencia**. (Estas tablas han sido preparadas por Cohen, 1988, y Kraemer & Thiemann, 1987, entre otros). En los capítulos siguientes, con cada método tratado daremos las tablas de potencia básicas y veremos cómo utilizarlas. En el apéndice B ofrecemos un índice de estas tablas bajo el nombre de tabla B-5.

La lógica en la que se basan estas tablas es precisamente lo que hemos aprendido aquí, y utilizar las tablas requiere exactamente la misma información que el cálculo directo de la potencia. De todos modos, el objetivo de este capítulo es ayudar a comprender el concepto de potencia, y x = 58

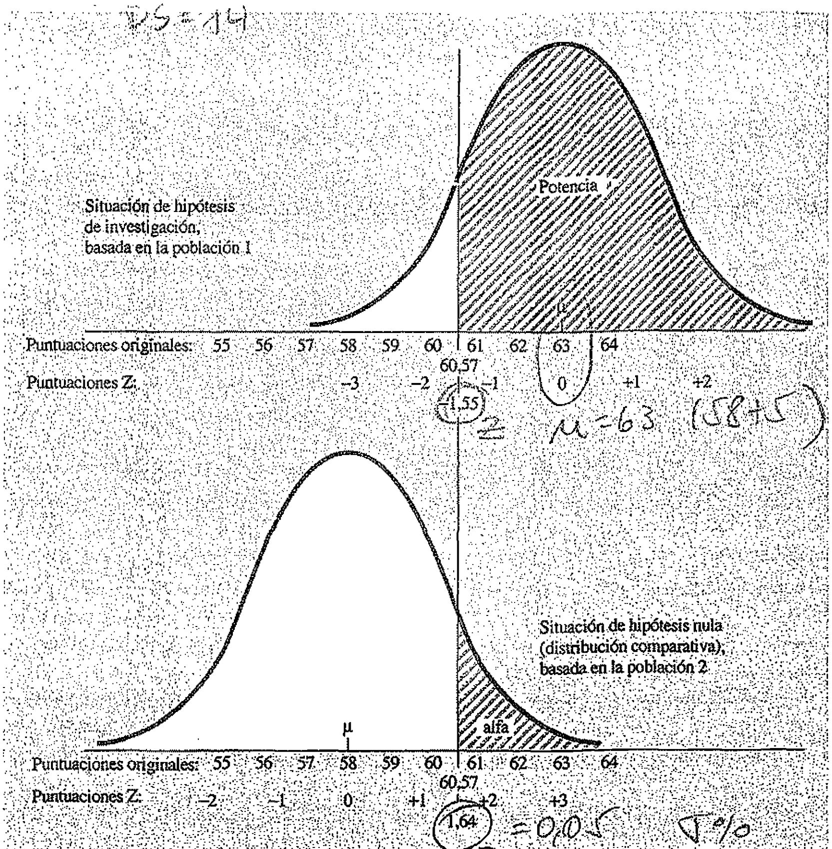


Figura 8-3. Distribuciones de medias de 80 observaciones basadas en distribuciones predichas (superior) y conocidas (inferior) de poblaciones de un estudio ficticio de empleados que reciben un nuevo programa de promoción sanitaria. En ambas distribuciones se indican las puntuaciones Z y las puntuaciones originales correspondientes al punto de corte de la distribución inferior (el punto de corte corresponde a un nivel de significación de $p<0,05$, prueba de una cola).

no sólo calcular el número. Es especialmente importante comprender qué factores influyen en la potencia y cómo se aplica todo esto a la planificación de experimentos y a la interpretación de los resultados de las investigaciones.

¿QUÉ FACTORES DETERMINAN LA POTENCIA DE UN ESTUDIO?

La potencia de un estudio depende de dos factores principales. El primero es el tamaño del efecto predicho por la hipótesis de investigación (tamaño de efecto). El segundo factor principal es la cantidad de participantes que incluye el estudio (tamaño de la muestra). La potencia también se ve afectada por a) el nivel de significación elegido, b) si se utiliza una prueba de una o dos colas, y c) el tipo de procedimiento de prueba de hipótesis utilizado.

TAMAÑO DEL EFECTO

Analicemos nuevamente el ejemplo de los investigadores expertos en educación que estudian el efecto de instrucciones especiales en alumnos de quinto grado que rinden un examen estándar para la evaluación de nivel. Las figuras 8-1 y 8-2 reflejan la situación en la que los investigadores predijeron que aquellos que recibían instrucciones especiales (población 1, curva superior) tendrían una media ocho puntos más alta que la de alumnos de quinto grado en general (población 2). La figura 8-4 se refiere al mismo estudio. Sin embargo, refleja una situación en la que los investigadores predicen que aquellos que recibieron las instrucciones especiales obtendrían una media 16 puntos más alta que la de los alumnos de quinto grado en general. Comparando la figura 8-4 con la figura 8-2, podemos notar que existe mayor probabilidad de obtener un resultado significativo si la situación reflejada en la figura 8-4 es verdadera. Lo que sucede es que hay más superposición de la curva superior con el área sombreada en la distribución comparativa.

Anteriormente calculamos que la probabilidad de obtener un resultado significativo (potencia), en el caso de la situación reflejada en las figuras 8-1 y 8-2, es de sólo el 38%. Si hiciéramos el mismo cálculo para el caso de la situación reflejada en la figura 8-4, descubriríamos que la potencia es del 85%. (Tal vez al alumno le interese probar la veracidad de este dato). La idea general es que cuanto mayor es la diferencia esperada entre las medias de las dos poblaciones, mayor potencia tiene el estudio.

La figura 8-5 ilustra dos distribuciones de medias basándose en el mismo ejemplo. Sin embargo, esta vez hemos cambiado el ejemplo de manera tal que la varianza sea mucho menor (el desvío estándar en la distribución de medias es exactamente la mitad de lo que era en las figuras 8-1, 8-2 y 8-4). En esta versión, la media predicha es la original de 208 (de las figuras 8-1 y 8-2). Sin embargo, ambas distribuciones de medias son mucho más estrechas, por lo tanto, hay mucha menos superposición entre la curva superior y la inferior (la distribución comparativa). El resultado es una potencia del 85%, mucho mayor que la de la situación original en las figuras 8-1 y 8-2. La idea, en este caso, es que, a menor varianza, mayor potencia.[4]

En conjunto, estos ejemplos ilustran el principio general que establece que, a menor superposición entre las dos distribuciones, mayor será la probabilidad de que el estudio arroje un resultado significativo. Dos distribuciones pueden tener poca superposición, tanto por una gran diferencia entre sus medias (figura 8-4) como por tener tan poca varianza, que incluso con una pequeña diferencia entre medias no se superponen demasiado (figura 8-5). La figura 8-6 resume este principio de modo más general.

La medida en la cual dos distribuciones no se superponen se denomina **tamaño del efecto**, porque es la medida en la cual el experimento tiene el efecto de separar las dos poblaciones. Es decir, cuanto mayor es la diferencia esperada entre las medias de las dos poblaciones, mayor es el tamaño del efecto, y a menor varianza en las dos poblaciones, mayor es el tamaño del efecto. En cualquier caso, a mayor tamaño de efecto, mayor potencia.

[4] Tal vez haya resultado evidente para el alumno que aumentamos la potencia a exactamente el 85%, tanto al duplicar el aumento predicho de las medias (como en la figura 8-4) o al reducir el desvío estándar de la distribución de medias a la mitad (como en la figura 8-5). Pronto veremos las razones por las cuales cualquiera de estos dos cambios produce el mismo resultado.

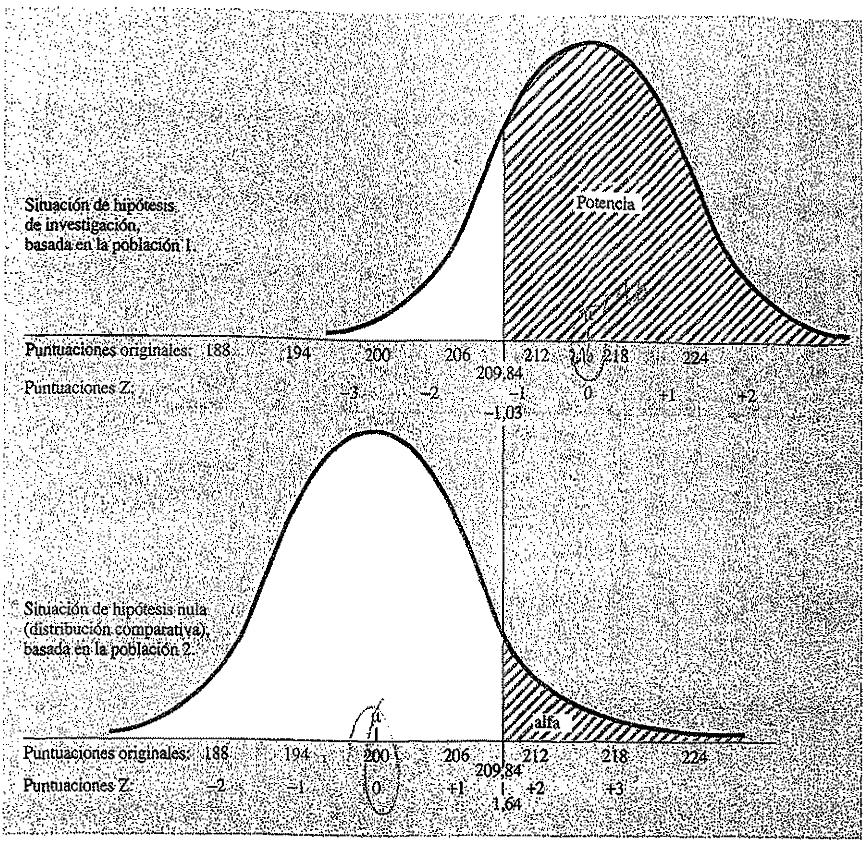


Figura 8-4. Distribuciones de medias de 64 resultados de exámenes basadas en distribuciones predichas (curva superior) y conocidas (curva inferior) de un estudio ficticio de alumnos de quinto grado que reciben instrucciones especiales antes de rendir un examen estándar para evaluar el nivel. En ambas distribuciones se indican las puntuaciones Z y los puntos de corte basados en la distribución inferior. (El punto de corte corresponde a un nivel de significación de $p < 0{,}05$, prueba de una cola). En este ejemplo, la media predicha de la distribución superior es 216.

T° (= dif medias)

Cálculo del tamaño del efecto

DS Pob?

Al determinar la potencia antes de realizar el estudio, el tamaño del efecto se calcula sobre la base de dos números. El primer número es la predicción del investigador en cuanto a la diferencia entre las medias de las dos poblaciones. La predicción se realiza sobre la base de determinada teoría, de experiencia previa en investigaciones de este tipo, o de lo que sería la menor diferencia ~~útil.~~ El segundo número es el desvío estándar poblacional. En los casos que hemos analizado hasta ahora, el desvío estándar (o la varianza) se conoce con anterioridad. (En capítulos posteriores analizaremos modos de estimar este dato cuando no se lo conoce).

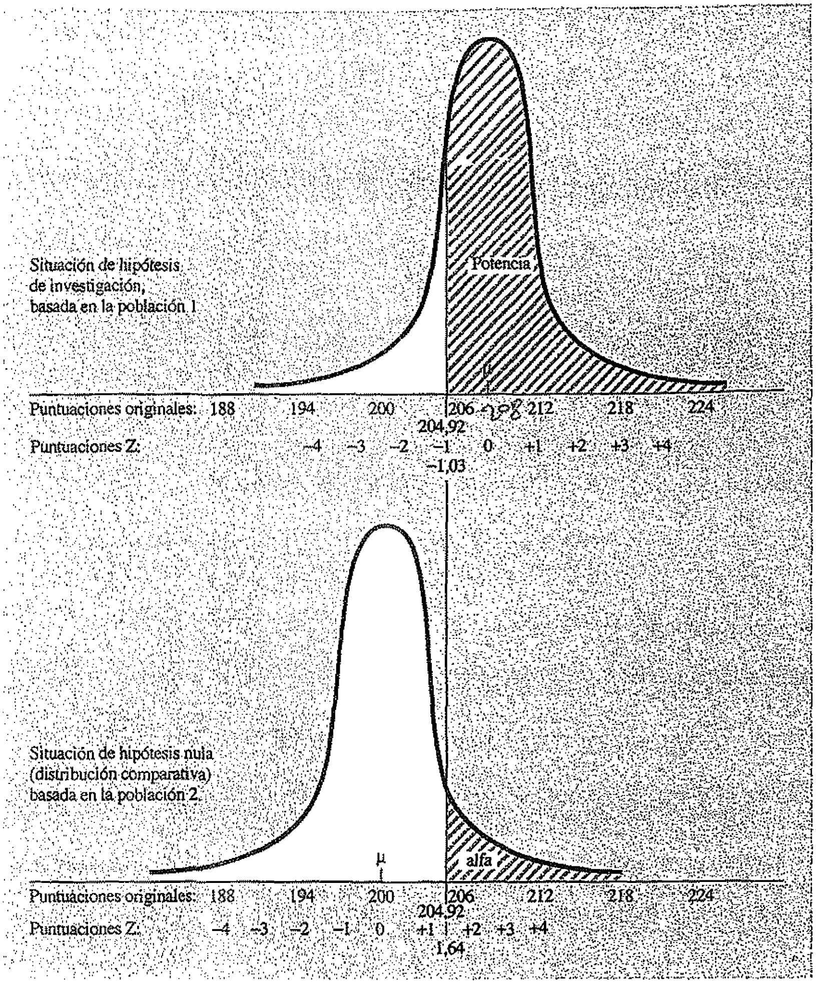


Figura 8-5. Distribuciones de medias de 64 resultados de exámenes basadas en distribuciones predichas (curva superior) y conocidas (curva inferior) de un estudio ficticio de alumnos de quinto grado que reciben instrucciones especiales antes de rendir un examen estándar para evaluar el nivel. En ambas distribuciones se indican las puntuaciones Z y los puntos de corte basados en la distribución inferior. (El punto de corte corresponde a un nivel de significación de $p < 0{,}05$, prueba de una cola). En este ejemplo, la media predicha de la distribución superior es ~~216~~.
206

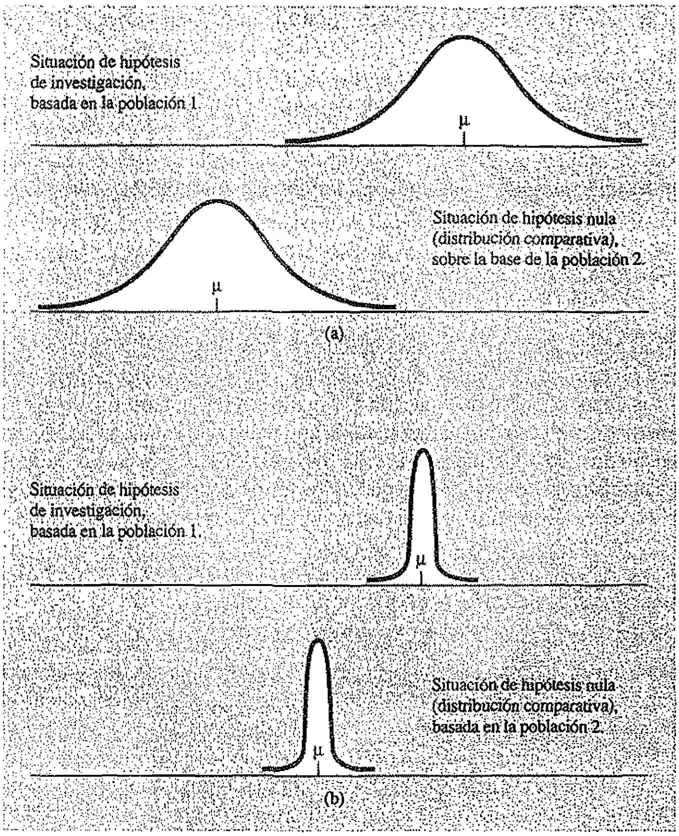


Figura 8-6. Las distribuciones de medias predichas y comparativas podrían tener poca superposición (y en ese caso el estudio tendría una potencia alta) debido a que (a) las dos medias son muy diferentes o (b) la varianza es pequeña.

La regla para el cálculo del tamaño del efecto es la siguiente: dividir la diferencia predicha entre las medias por el desvío estándar poblacional. La fórmula sería la siguiente:

$$d = \frac{\mu_1 - \mu_2}{\sigma} \tag{8-1}$$

En esta fórmula, d es el símbolo del tamaño del efecto (también conocido como "d de Cohen". (En capítulos posteriores veremos otras medidas del tamaño de efecto apropiadas para diferentes situaciones, que se representan con otros símbolos). μ_1 es la media de la población 1 (la media predicha para la población que recibe la manipulación experimental); μ_2 es la media de la población 2 (distribución comparativa), y σ es el desvío estándar de la población 2. Es importante tener en cuenta que al calcular el tamaño del efecto, no utilizamos el desvío estándar de la distribución de medias (σ_M). En cambio, utilizamos el desvío estándar de la población de observaciones originales (σ). (Cabe mencionar también que sólo nos interesa el σ de una población, ya que en la prueba de hipótesis generalmente suponemos que ambas poblaciones tienen el mismo desvío estándar. Más adelante volveremos a tratar este tema).

En el primer ejemplo de este capítulo (figuras 8-1 y 8-2), la diferencia entre las medias era 8 y el desvío estándar de la población original de individuos, alumnos de quinto grado, era 48. Por lo tanto, el tamaño del efecto era 8/48, es decir, 0,17. La fórmula sería la siguiente:

$$d = \frac{\mu_1 - \mu_2}{\sigma} = \frac{208 - 200}{48} = \frac{8}{48} = 0{,}17$$

Analicemos ahora el ejemplo en el que la diferencia de medias era de 16 puntos de examen y el desvío estándar poblacional también era 48 (figura 8-4). En ese caso, el tamaño del efecto es 0,33 (16/48 = 0,33), es decir, el doble del anterior. De modo similar, analicemos el ejemplo en el que la diferencia de medias era 8 pero con una población con un desvío estándar de 24 (figura 8-5). En ese ejemplo, el tamaño de efecto también es de 0,33 (es decir, 8/24 = 0,33). (La potencia en los dos casos también era la misma –85%–, debido a que todos los otros aspectos del estudio eran los mismos. Por lo tanto, si tienen el mismo tamaño del efecto, tendrán la misma potencia).

Un alcance más general del tamaño del efecto

El tamaño del efecto, como hemos visto, es la diferencia de medias dividida por el desvío estándar poblacional. Esta división estandariza la diferencia entre las medias y ubica la diferencia en una escala adaptada al desvío estándar de la medida utilizada. Este proceso tiene el mismo tipo de efecto que convertir una puntuación original en una puntuación Z. En ambos casos, el resultado es una base estándar de comparación con otros valores, incluso valores de diferentes escalas. Supongamos que dos estudios utilizan diferentes medidas. (Por ejemplo, un grupo de investigadores expertos en educación estudia el efecto de instrucciones especiales utilizando la "Prueba de niveles académicos de Jones" con un $\sigma = 48$, y otro grupo de investigadores utiliza la "Prueba de logros académicos de Smith" con un $\sigma = 17$). Incluso en esta situación, dado que las diferencias de medias están divididas por el desvío estándar, se puede comparar directamente el tamaño del efecto de los dos casos.

La estandarización que proporciona el tamaño del efecto (*d*) es especialmente útil porque se basa en el desvío estándar de la población de observaciones individuales (en lugar del desvío estándar de la distribución de medias). Esto significa que podemos utilizar *d* para comparar resultados de estudios muy diferentes, incluso de aquellos que utilizan diferentes tamaños de muestras.

En resumen, supongamos que un estudio tiene un tamaño de efecto (*d*) de 0,25. Esto siempre significa que existe un cuarto de desvío estándar de diferencia entre las dos medias, independientemente del tamaño de la muestra y de la medida utilizada. Si un estudio tiene una *d* de 0,25, y otro una *d* de 2,0 (diferencia de 2 desvíos estándar entre las medias), sabríamos que el efecto fue mucho mayor en el segundo estudio, aun si las medidas utilizadas y la cantidad de participantes en los dos estudios fueran completamente diferentes. (Más adelante en el capítulo veremos que una aplicación importante del tamaño del efecto se utiliza en el procedimiento denominado meta-análisis, el cual proporciona a los investigadores una herramienta precisa y objetiva para utilizar el tamaño del efecto, con el fin de combinar y comparar los resultados de distintos estudios acumulados acerca de un tema determinado, como por ejemplo, la utilidad de determinado tipo de psicoterapia o la diferencia ente dos grupos de edades con respecto a una capacidad).

Reglas del tamaño del efecto

difícil de conocer antes del estudio

Es difícil saber, antes de realizar un estudio, cuál es el tamaño del efecto que debemos esperar; si lo supiéramos, no necesitaríamos realizar la investigación. Jacob Cohen (1988, 1992), un psicólogo que ha trabajado mucho en lo que se refiere al desarrollo de los cálculos estadísticos relacionados con la potencia, ha colaborado en la solución de este problema. Cohen ha establecido algunas **reglas del tamaño del efecto** basadas en los efectos descubiertos a través de las investigaciones psicológicas en general. Estas reglas, al menos, indican al investigador cuándo considerar que un efecto es pequeño, mediano o grande. Así, si el investigador cree que determinado estudio debería tener un efecto mediano, ahora cuenta con un número específico que, según ha descubierto Cohen, es típico de los efectos medianos, y que puede entonces utilizar para calcular la potencia.

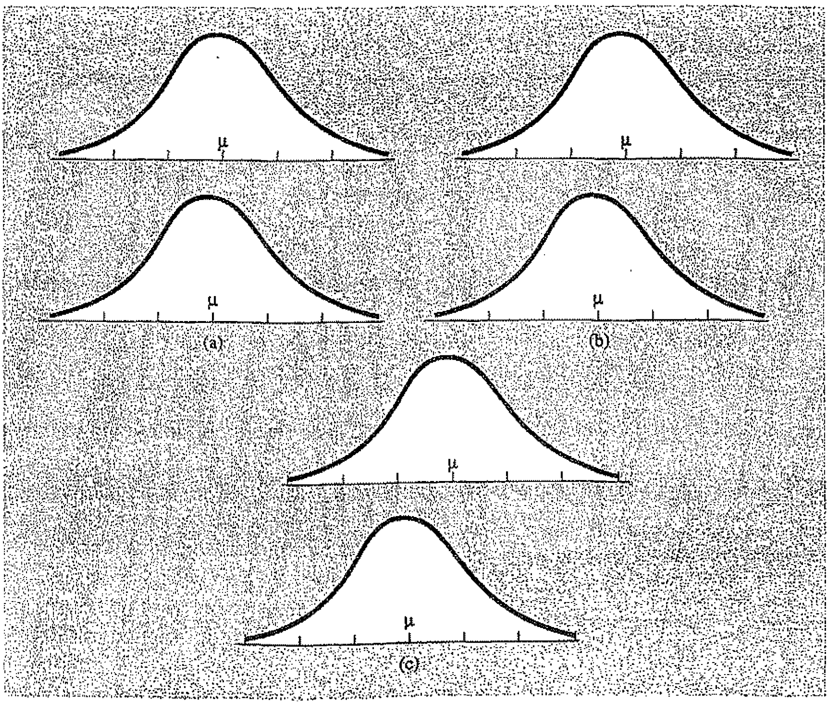


Figura 8-7. Comparación de pares de distribuciones poblacionales de individuos que indican las reglas de Cohen para el tamaño del efecto: (a) tamaño del efecto pequeño ($d = 0,2$), (b) tamaño del efecto mediano ($d = 0,5$) y (c) gran tamaño del efecto ($d = 0,8$).

Recordemos que hemos calculado el tamaño del efecto como la diferencia predicha entre las medias de las dos poblaciones dividida por el desvío estándar poblacional. Cohen recomienda que para el tipo de situación que estamos analizando en este capítulo, deberíamos pensar en un "pequeño tamaño del efecto", aproximadamente 0,20. Con una *d* de 0,20, las poblaciones de observaciones individuales tienen una superposición de aproximadamente un 85%. Ésta es el tamaño del efecto de, por ejemplo, la diferencia de altura entre niñas de 15 y 16 años de edad (*véase* figura 8-7a), que es de aproximadamente 1/2 pulgada de diferencia, con un desvío estándar de aproximadamente 2,1 pulgadas. (Cuando en estos ejemplos hablamos de porcentaje de superposición, nos referimos a la superposición de las poblaciones de observaciones individuales. La cantidad de superposición de las distribuciones de medias será menor según el tamaño de la muestra).

Cohen considera que un tamaño del efecto mediano es de 0,5, el cual implica una superposición de aproximadamente el 67%, que es aproximadamente la diferencia de altura entre niñas de 14 y 18 años de edad (*véase* figura 8-7b). Finalmente, Cohen define un gran tamaño del efecto en 0,80. Esto implica una superposición de sólo un 53%, que es aproximadamente la diferencia de altura entre niñas de 13 y 18 años de edad (*véase* figura 8-7c). Las tres reglas del tamaño del efecto que acabamos de mencionar están resumidas en la tabla 8-2.

Analicemos otro ejemplo. Como observamos anteriormente en este libro, muchas pruebas de CI tienen un desvío estándar de 16 puntos. Un procedimiento experimental diseñado para aumentar el CI, que tuviera un pequeño tamaño del efecto, implicaría un aumento del CI de 3,2 puntos CI. (Una diferencia de 3,2 puntos CI, entre la media de la población que recibió el procedimiento experimental y la media de la población que no lo recibió, dividida por el desvío estándar de la población, que es de 16, arroja un tamaño del efecto de 0,20, es decir, $d = 3{,}2/16 = 0{,}20$). Un procedimiento experimental con un tamaño del efecto mediano aumentaría el CI en 8 puntos. Un procedimiento experimental con un gran tamaño del efecto aumentaría el CI en 12,8 puntos. Para dar otro ejemplo, analicemos los registros del SAT (σ es aproximadamente 100). En una población con una puntuación media de SAT igual a 500, los individuos que participaran de un procedimiento experimental para aumentar los valores del SAT con un pequeño tamaño del efecto lograrían un valor medio de 520; aquellos que participaran de un procedimiento con un tamaño del efecto mediano lograrían un valor medio de 550, y aquellos que participaran de un procedimiento con un gran tamaño del efecto lograrían un valor medio de 580. Finalmente, la figura 8-8 representa gráficamente tamaños del efecto pequeños y medianos aplicados a las distribuciones de medias del ejemplo de los alumnos de quinto grado. Podemos observar que cuanto mayor es el tamaño del efecto, mayor es la potencia.

Las reglas del tamaño del efecto establecidas por Cohen son importantes para los científicos, porque en la mayoría de las investigaciones es difícil saber de antemano qué tamaño del efecto predecir. (Si uno no puede predecir el tamaño del efecto, no puede siquiera buscar la potencia en la tabla). Algunas veces, los investigadores pueden basar sus predicciones en cuanto al tamaño del efecto en investigaciones o teorías previas. *Asimismo, también existe un mínimo tamaño* del efecto que sería importante para algún objetivo práctico. Pero en la mayoría de los casos, los investigadores están analizando un tema por primera vez, y sólo tienen una idea vaga del tamaño del efecto a esperar. Las reglas de Cohen ayudan a los investigadores a convertir esa vaga idea en un número.

Tabla 8-2.
Resumen de las reglas de Cohen del tamaño del efecto para las diferencias de medias.

Descripción verbal	Tamaño del efecto (*d*)
Pequeño	0,20
Mediano	0,50
Grande	0,80

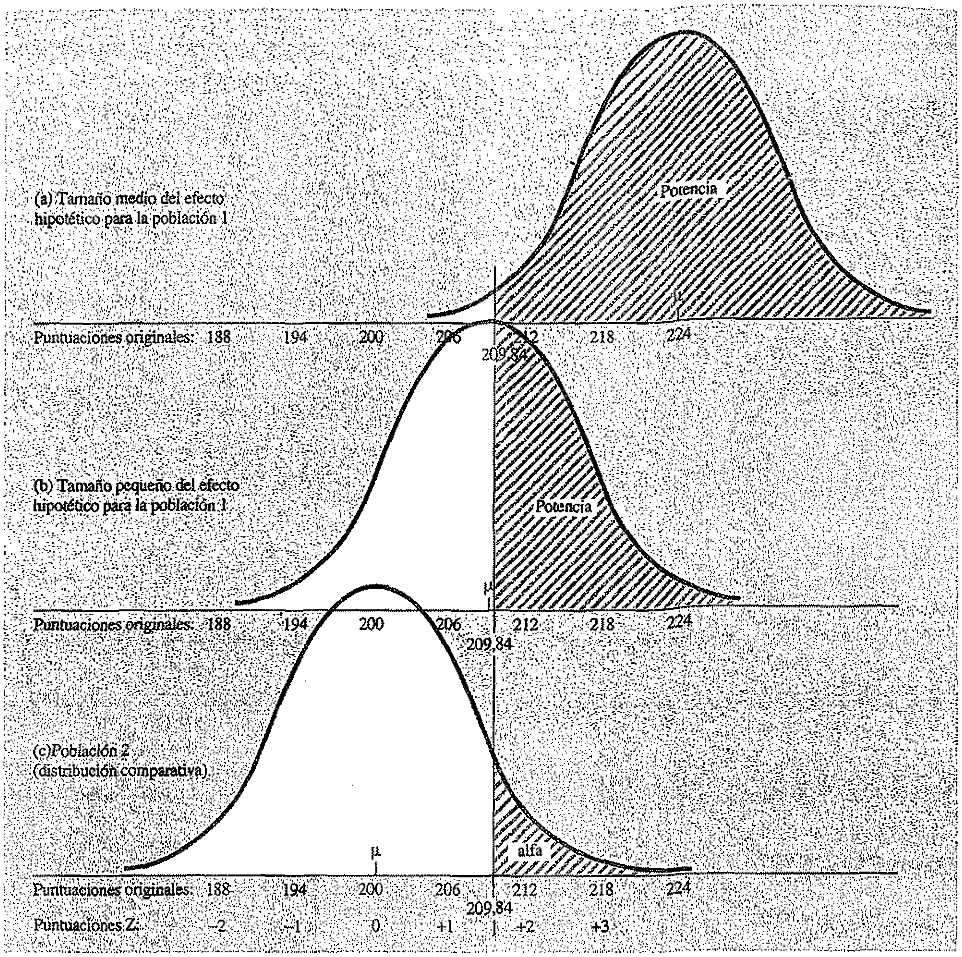


Figura 8-8. Distribuciones de poblaciones de medias predichas (a, b) y conocidas (c) en un estudio ficticio de alumnos de quinto grado que reciben instrucciones especiales antes de rendir un examen estándar para la evaluación de nivel. En las distribuciones a, b y c se indican las puntuaciones originales correspondientes a la puntuación Z de corte que se muestra en la distribución c. (El punto de corte está calculado sobre la base de un nivel de significación $p < 0{,}05$, prueba de una cola). En este ejemplo (a) es la distribución predicha con un tamaño mediano del efecto ($d = 0{,}5$, potencia = 99%), y (b) es la distribución predicha con un tamaño pequeño del efecto ($d = 0{,}2$, potencia = 48%).

Las reglas de Cohen para el tamaño del efecto también son útiles para interpretar resultados de estudios. Nos proporcionan un parámetro para decidir acerca de la importancia del efecto de un estudio con relación a lo que es típico en psicología.

TAMAÑO DE LA MUESTRA

El otro factor de influencia importante en la potencia, además del tamaño del efecto, es la cantidad de personas que integran la muestra estudiada. Básicamente, a mayor cantidad de personas, mayor potencia.

El tamaño de la muestra influye en la potencia porque, a mayor tamaño de muestra, menor es el desvío estándar de la distribución de medias. Si las distribuciones tienen un desvío estándar menor, son más estrechas y, por ende, están menos superpuestas. La figura 8-9 representa gráficamente la situación que se plantearía en el ejemplo de alumnos de quinto grado si el estudio incluyera 100 alumnos en lugar de los 64 del ejemplo original (figuras 8-1 y 8-2). La potencia en este caso es del 51% (con 64 alumnos era del 38%). Con un estudio de 500 participantes, la potencia es del 98% (*véase* figura 8-10).

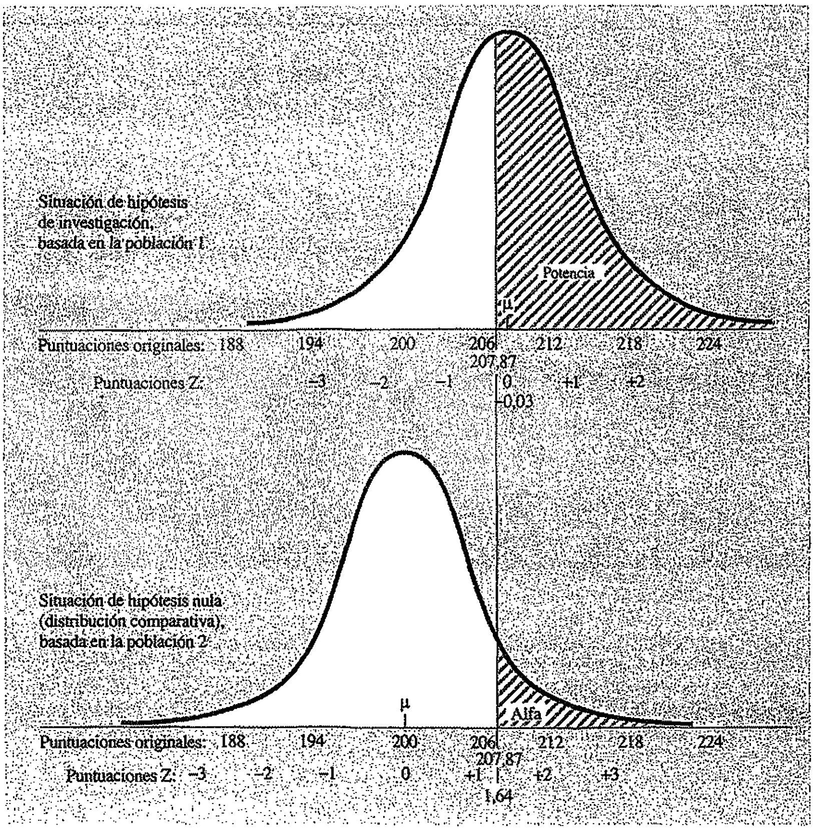


Figura 8-9. Distribuciones de medias de 100 resultados de exámenes (en lugar de 64, como en las figuras 8-1 y 8-2) basadas en distribuciones predichas (superior) y conocidas (inferior) de un estudio ficticio realizado a alumnos de quinto grado que reciben instrucciones especiales antes de rendir un examen estándar para la evaluación de nivel. En las dos distribuciones se indican los puntos de corte según la distribución inferior. (El punto de corte corresponde a un nivel de significación de $p < 0,05$, prueba de una cola). Potencia 51%.

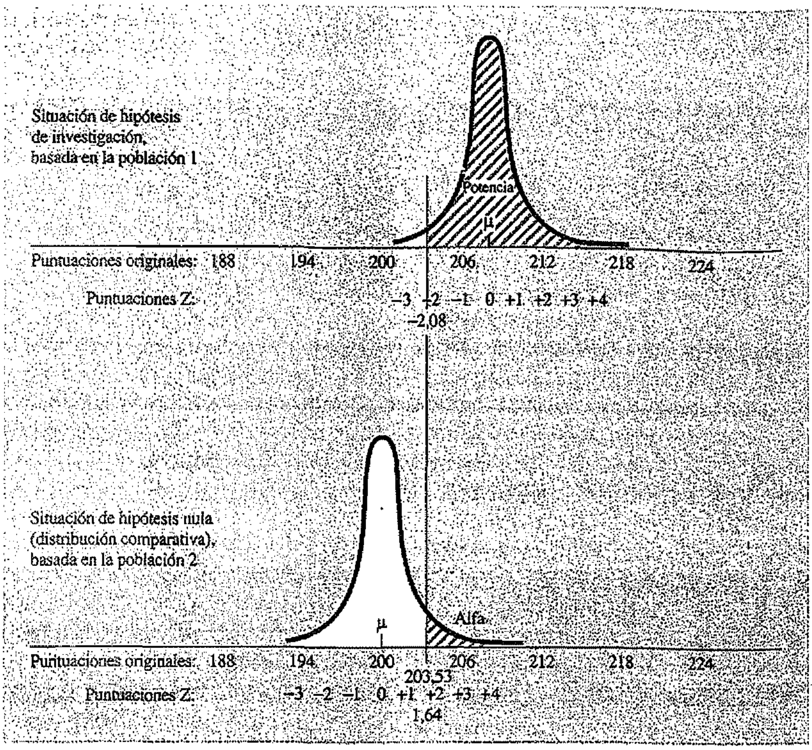


Figura 8-10. Distribuciones de medias de 500 resultados de exámenes, basadas en distribuciones predichas (superior) y conocidas (inferior) de un estudio ficticio realizado a alumnos de quinto grado que reciben instrucciones especiales antes de rendir un examen estándar de evaluación de nivel. En las dos distribuciones se indican los puntos de corte según la distribución inferior. (El punto de corte corresponde a un nivel de significación $p < 0{,}05$, prueba de una cola). Potencia 99%.

No debemos confundirnos. Las distribuciones de medias pueden ser estrechas (y por lo tanto estar menos superpuestas y tener más potencia) por dos razones muy diferentes. Una razón es que las poblaciones de individuos pueden tener desvíos estándar pequeños. Este motivo está relacionado con el tamaño de efecto. La otra razón por la que las dos distribuciones de medias pueden ser estrechas es que el tamaño de la muestra sea grande. Este motivo es completamente independiente del primero. El tamaño de la muestra no tiene nada que ver con el tamaño del efecto, y tanto el primero como el segundo influyen en la potencia. Pero como veremos pronto, estas dos influencias distintas sobre la potencia llevan a pasos prácticos completamente diferentes para aumentar la potencia al planificar un estudio.

Cálculo del tamaño de muestra necesario para determinado nivel de potencia

La razón principal por la que los investigadores calculan la potencia al planificar un estudio es para decidir cuántos participantes incluir en el mismo. Dado que el tamaño de la muestra es un factor de influencia importante en la potencia, los investigadores necesitan estar seguros de tener suficientes participantes como para que sus estudios tengan un nivel de potencia bastante alto.

Un investigador puede calcular la cantidad necesaria de participantes revirtiendo los pasos para el cálculo de la potencia. Comenzamos con el nivel de potencia deseado, digamos, un 80%, y luego calculamos cuántos participantes necesitaríamos para obtener ese nivel de potencia. Supongamos que los psicólogos especializados en educación, quienes realizaron el ejemplo de los alumnos de quinto grado, estuvieran planificando ese estudio y quisieran calcular cuántos alumnos de quinto grado necesitan analizar. Siendo la diferencia de media predicha igual a 8, y el desvío estándar de la población conocida igual a 48, necesitarían 222 alumnos de quinto grado para tener una potencia del 80%. En este momento no entraremos en detalles de cálculo. (Sin embargo, el alumno tal vez quiera intentar calcular este dato por sí mismo. Sería interesante ver si puede llegar a la misma respuesta que nosotros utilizando los procedimientos que ha aprendido, pero comenzando con una potencia del 80% y continuando con los pasos desde atrás hacia adelante para obtener la cantidad necesaria de participantes). En la práctica, los investigadores utilizan tablas especiales que especifican cuántos participantes son necesarios en un estudio para tener un alto nivel de potencia, según un determinado tamaño del efecto. Nosotros proporcionaremos versiones simplificadas de esas tablas para cada uno de los principales procedimientos de prueba de hipótesis que veremos en los capítulos siguientes.

Cuadro 8-1. La potencia de los experimentos psicológicos típicos.

Hace más de tres décadas, Jacob Cohen (1962), un psicólogo especialista en métodos estadísticos, publicó un análisis, muy conocido actualmente de la potencia estadística de estudios publicados en el volúmen de 1960 de la *Revista Científica de Psicología Patológica y Social [Journal of Abnormal and Social Psychology]*. Cohen observó que se prestaba gran importancia a la significación, o también a si se había cometido un error Tipo I (es decir, si se había rechazado equivocadamente la hipótesis nula y a partir de los resultados se había supuesto cierto efecto que en realidad no existía). Pero esencialmente no se prestaba atención a la posibilidad de un error Tipo II (es decir, si por error no se hubiera rechazado la hipótesis nula y se hubiera ignorado un efecto real debido a resultados no concluyentes, que de hecho algunas veces eran considerados inexistentes). En estos estudios la potencia ni siquiera se discutía.

Cohen calculó la potencia de los resultados de esas publicaciones. Al no estar familiarizado con muchos de los contenidos de las distintas áreas, analizó la potencia según tres supuestos del tamaño del efecto: pequeño, mediano y grande. Descubrió que si era pequeña, los estudios publicados tenían sólo una chance contra seis de detectar algún efecto. *Ninguno tenía mayores chances* que un 50%. Si suponía un efecto mediano en la población, los estudios tenían chances apenas mayores a un 8% de detectar ese efecto, e incluso un cuarto de ellos tenía menos de una chance contra tres. Sólo el supuesto de grandes efectos daba a los estudios, tal como estaban diseñados, una buena posibilidad de rechazar la hipótesis nula. Como el mismo Cohen lo expresara "toda una generación de investigadores podría gozar de un empleo adecuado si se intentaran repetir estudios interesantes que originalmente utilizaron tamaños de muestras inadecuados. Lamentablemente, aquellos que más merecerían tales repeticiones son los que seguramente no han sido impresos". (p. 153)

En la última oración, Cohen se refiere al hecho de que muchos más estudios que hubieran sido apropiados para la revista científica mencionada, probablemente nunca se escribieron porque los investigadores obtuvieron resultados no significativos, que las probabilidades claramente indicaban que estaban predestinados a obtener dado el bajo nivel de potencia de la mayoría de los estudios en ese campo. Los experimentos que "fallaron", cuando en realidad sus hipótesis nunca fueron adecuadamente probadas, representaban una gran cantidad de conocimiento que posiblemente se haya perdido, y que tal vez nunca se vuelva a investigar. Y esa pérdida se debió simplemente a la falta de interés en la potencia, en la mayoría de los casos por la falta de cálculo (a través del análisis del tamaño del efecto, el nivel de significación y la potencia) del tamaño óptimo de la muestra para probar la hipótesis de la mejor manera posible.

Después del análisis realizado por Cohen, se han llevado a cabo varios análisis similares acerca de la potencia, publicados en determinadas revistas científicas (p. ej. el análisis de Brewer de 1972 en la *Revista Científica Americana de Investigación Educativa [American Educational Research Journal]* y el estudio de Chase & Chase de 1976 en la *Revista Científica de Psicología Aplicada [Journal of Applied Psychology])*. Mientras tanto, en 1969 Cohen publicó una guía para el análisis de la potencia en las ciencias sociales, y una versión revisada apareció en 1988. Aun así, en una publicación de 1989, Sedlmeier y Gigerenzer observaron que las advertencias de Cohen aparentemente no tuvieron efecto durante los años siguientes. De hecho, la potencia de los estudios que aparecieron en la misma revista científica que Cohen había analizado (ahora llamada *Revista Científica de Psicología Patológica [Journal of Abnormal Psychology])*, en realidad había disminuido con el correr de esos años, y el bajo nivel de potencia continuaba pasando inadvertido. Sólo dos de sesenta y cuatro experimentos trataban el tema de la potencia, y en estos dos no se la había estimado.

Mientras tanto, en el 11% de los estudios publicados en esa edición, la falta de significación era tomada como una confirmación de la hipótesis nula, tal vez en un intento de adherir a las admoniciones tradicionales que cuestionamos en el cuadro 6-1. Sin embargo, Sedlmeier y Gigerenzer descubrieron que la potencia media en esos estudios era sólo de 0,25. Ciertamente, si cuando los resultados favorecen la hipótesis nula vamos a tratarlos como información valiosa en sí misma (nuevamente, *véase* cuadro 6-1), sólo podremos hacerlo si la potencia es lo suficientemente alta como para que, si la hipótesis de investigación fuera verdadera, el estudio tuviera al menos iguales chances de reflejarlo.

Esta obstinada omisión por parte de los investigadores es un poco sorprendente. La mayoría de las veces implica que realizan todo su trabajo para nada. Aunque lo que intenten demostrar sea cierto, tienen pocas probabilidades de lograrlo. Y aparentemente la metodología en psicología es tan monolítica y fija que no puede modificarse. Sin embargo, en una publicación en *Psicólogo Americano [American Psychologist]* titulada "Cosas que he aprendido (hasta ahora)", Jacob Cohen (1990) recuerda las décadas anteriores desde un punto de vista filosófico:

> No me desespero. Recuerdo que W. S. Gosset, el muchacho que trabajaba en una fábrica de cerveza y que publicó modestamente su trabajo como "El estudiante", publicó la prueba *t* una

década antes de que empezara la Primera Guerra Mundial, y esa prueba no formó parte de los libros de estadística aplicada a la psicología sino hasta después de la Segunda Guerra Mundial. Estas cosas llevan tiempo. Por lo tanto, si el alumno llegara a publicar algo que considera realmente bueno, y transcurre un año o una década o dos, y casi nadie parece prestarle atención, debe recordar la prueba *t* y tener confianza. (p. 1311)

OTROS FACTORES QUE INFLUYEN SOBRE LA POTENCIA

Existen otros tres factores (además del tamaño del efecto y del tamaño de la muestra) que afectan la potencia:

1. **Nivel de significación (alfa).** Un nivel de significación menos extremo (como 0,10) implica más potencia, y un nivel de significación más extremo (0,01 o 0,001) implica menos potencia. Un nivel menos extremo de significación da por resultado mayor potencia, porque cuando el nivel de significación no es muy extremo (como por ejemplo 0,10), el área de rechazo sombreada bajo la curva inferior es mayor. Por lo tanto, una mayor parte del área, bajo la curva superior, está sombreada. Un nivel de significación más extremo da por resultado menor potencia, porque cuando este nivel es más extremo (como por ejemplo 0,01), el área de rechazo sombreada bajo la curva inferior es menor. Supongamos que en nuestra versión original del ejemplo de los alumnos de quinto grado (figuras 8-1 y 8-2) hubiéramos utilizado el nivel de significación de 0,01 en lugar del de 0,05. La potencia hubiera caído del 38% a sólo el 16% (figura 8-11).
2. **Prueba de una cola versus prueba de dos colas.** Utilizar una prueba de dos colas hace que resulte más difícil obtener significación en cualquiera de las colas. Por lo tanto, si se mantienen iguales todas las demás condiciones, la potencia será menor con una prueba de dos colas que con una de una cola. Supongamos que hubiéramos utilizado una prueba de dos colas, en lugar de una, en nuestro ejemplo de los alumnos de quinto grado. Como lo ilustra la figura 8-12, la potencia sería sólo del 26% (comparado con el 38% de la versión original de una cola representada por las figuras 8-1 y 8-2).
3. **Tipo de procedimiento de prueba de hipótesis.** A veces, el investigador puede elegir entre más de un procedimiento de prueba de hipótesis para realizar determinado estudio. En este libro todavía no hemos analizado ninguna situación de este tipo, pero sí lo haremos en el capítulo 15.

Resumen de factores que influyen en la potencia

La tabla 8-3 resume el efecto de distintos factores en la potencia de un estudio.

PAPEL QUE DESEMPEÑA LA POTENCIA AL DISEÑAR UN EXPERIMENTO

Para planificar un estudio es crucial determinar su potencia. Cuando la potencia del estudio que se planifica es baja, aun si la hipótesis de investigación resultara verdadera, sería muy poco probable que el estudio arroje resultados significativos. De ese modo, se estaría desperdiciando el tiempo y dinero que implican realizar el estudio. Por lo tanto, cuando se descubre que la potencia de un estudio es baja, los investigadores intentan encontrar formas prácticas de aumentarlo hasta un nivel aceptable.

¿Qué significa un nivel aceptable de potencia? Cohen (1988) sugiere que, por lo general, un estudio debería tener aproximadamente un 80% de potencia para que valga la pena realizarlo. Obviamente, cuanto más potencia, mejor. Sin embargo, los costos que implica obtener más potencia (como por ejemplo analizar una mayor cantidad de personas) hacen, con frecuencia, que hasta una potencia del 80% esté fuera de nuestro alcance.

¿Cómo podemos aumentar la potencia de una investigación? En principio, la potencia de una investigación puede aumentarse cambiando cualquiera de los factores resumidos en la tabla 8-3. Analicemos cada uno de ellos.

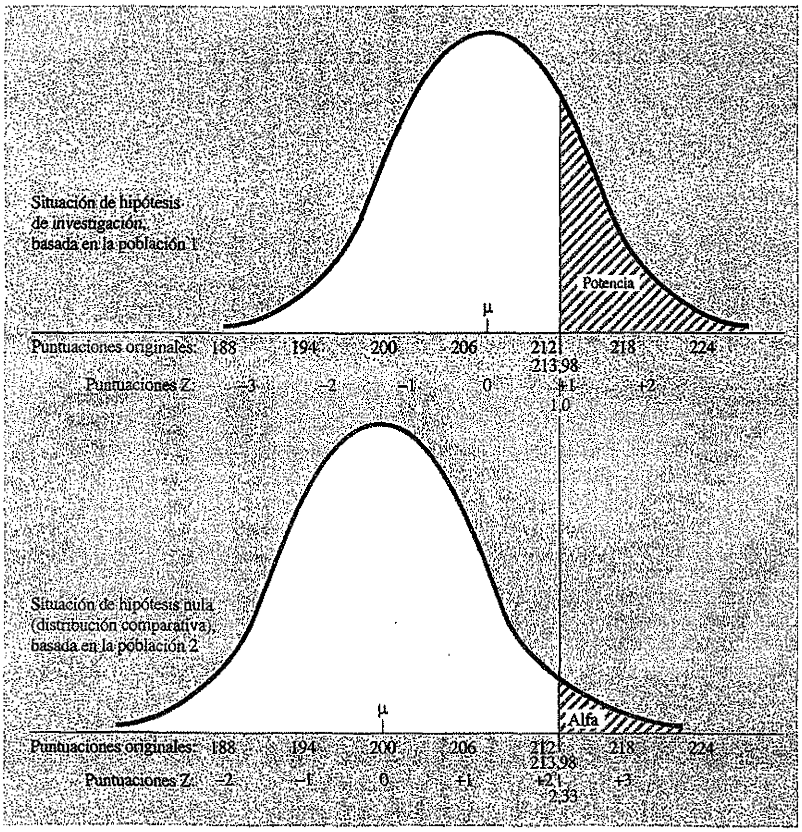


Figura 8-11. Distribuciones de medias de 64 resultados de exámenes basadas en distribuciones predichas (superior) y conocidas (inferior) de un estudio ficticio realizado con alumnos de quinto grado que reciben instrucciones especiales antes de rendir un examen estándar para la evaluación de nivel. En las dos distribuciones se indican las puntuaciones Z y originales de corte de la distribución inferior. El punto de corte correspondiente a un nivel de significación de $p < 0,01$, prueba de una cola (en comparación con el $p < 0,05$ del ejemplo original representado por las figuras 8-1 y 8-2). Potencia 16%.

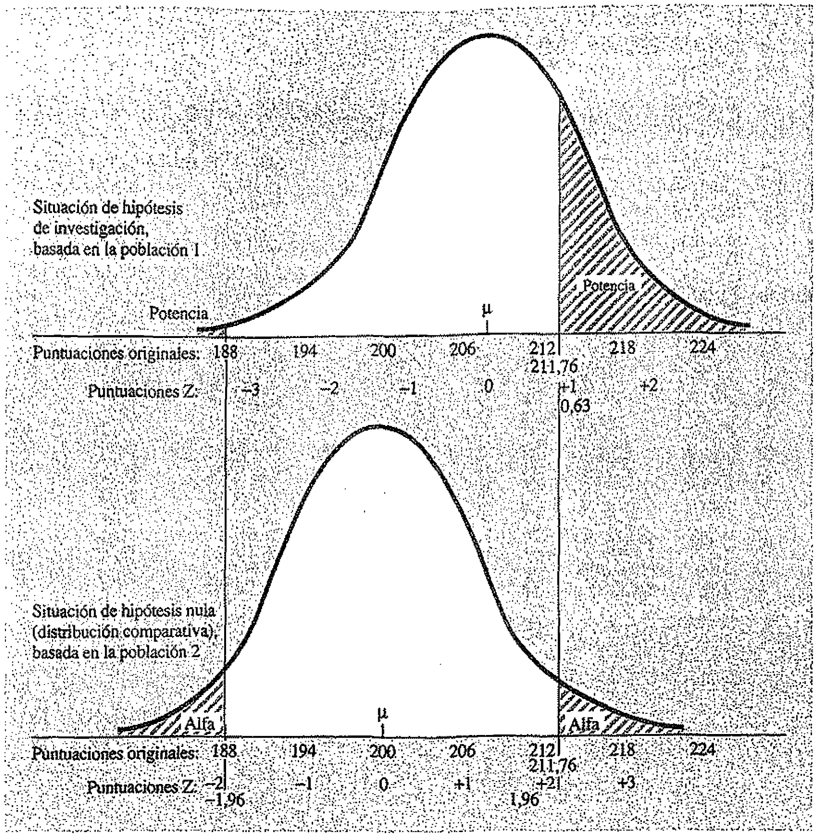


Figura 8-12. Distribuciones de medias de 64 resultados de exámenes basadas en distribuciones predichas (curva superior) y conocidas (curva inferior) de un estudio ficticio realizado a alumnos de quinto grado que reciben instrucciones especiales antes de rendir un examen estándar para la evaluación de nivel. En las dos distribuciones se indican las puntuaciones Z y originales de corte de la distribución inferior. Los puntos de corte corresponden a un nivel de significación de $p < 0,05$, prueba de dos colas (en comparación con la prueba de una cola del ejemplo original representado por las figuras 8-1 y 8-2). Potencia = 26%.

1. **Aumentar la diferencia predicha entre medias poblacionales.** Un investigador no puede simplemente aumentar en forma arbitraria su predicción en cuanto a esta diferencia. Si lo hiciera, aumentaría la potencia calculada, pero en realidad no lograría que el estudio tuviera más probabilidades de arrojar un resultado significativo. A veces, sin embargo, es posible cambiar el modo en el que se realiza el estudio para que el investigador tenga motivos para esperar una mayor diferencia de medias. Analicemos nuevamente el ejemplo basado en el experimento acerca del impacto producido por instrucciones especiales en los resultados de los exámenes realizados por alumnos de quinto grado. Una manera de aumentar la diferencia de medias esperada podría ser trabajar con instrucciones más elaboradas, explicarlas con más detenimiento, tal vez dar tiempo para que

Tabla 8-3.
Factores que influyen en la potencia.

Características del estudio	**Aumenta la potencia**	**Disminuye la potencia**
Tamaño del efecto (d) ($d = [\mu_1 - \mu_2]/\sigma$)	d Grande	d Pequeña
El tamaño del efecto combina las dos características siguientes:		
Diferencia hipotética entre medias poblacionales ($\mu_1 - \mu_2$)	Grandes diferencias	Pequeñas diferencias
Desvío estándar poblacional (σ)	σ pequeño	σ grande
Tamaño de muestra (N)	N grande	N pequeño
Nivel de significación (α)	α indulgente, alto (tal como 0,05 ó 0,10)	α riguroso, bajo (tal como 0,01 ó 0,001)
Prueba de una cola versus prueba de dos colas	Una cola	Dos colas
Tipo de procedimiento de prueba de hipótesis utilizado	Varía	Varía

las practiquen, y otros cambios por el estilo. Una desventaja de este método es que puede ser difícil o costoso; o bien, puede requerir la utilización de un procedimiento experimental que no es igual al procedimiento al cual deseamos que se apliquen los resultados del estudio.

2. **Disminuir el desvío estándar poblacional**. Existen al menos dos modos de disminuir el desvío estándar poblacional de un estudio. Un método es realizar el estudio utilizando una población menos diversa que la población que se planeó utilizar originalmente. En el ejemplo basado en el examen de evaluación de nivel realizado por alumnos de quinto grado, podríamos utilizar sólo alumnos de quinto grado de determinado sistema escolar suburbano. La desventaja es que los resultados se aplican sólo a esa población más específica.

Otro método para disminuir el desvío estándar poblacional es utilizando condiciones y medidas de prueba más precisas. Por ejemplo, realizar la prueba en una situación estandarizada o en un ambiente de laboratorio controlado produce generalmente una variación general menor entre las observaciones (lo cual tiene como resultado un menor desvío estándar). De manera similar, utilizar pruebas con instrucciones claras y procedimientos precisos, en cuanto al modo de realizar las respuestas, también reduce la variación. Si estos cambios resultan prácticos, son métodos excelentes para aumentar la potencia, aunque por lo general el estudio ya es de por sí lo más riguroso posible.

3. **Aumentar el tamaño de la muestra**. El método más directo para aumentar la potencia de un experimento es mediante el análisis de una mayor cantidad de personas. Naturalmente, si estamos analizando astronautas que caminaron por la luna, existe un límite para esa cantidad. Sin embargo, en las situaciones reales de investigación el tamaño de la muestra es el principal método para modificar un estudio con el fin de obtener suficiente potencia.

4. **Utilizar un nivel de significación menos riguroso**. Comúnmente, el nivel de significación debería ser bastante riguroso de manera que proteja razonablemente el estudio del error Tipo I. Normalmente, este nivel será de 0,05. Por lo tanto, es raro que se pueda hacer mucho para aumentar la potencia en este sentido.

5. **Utilizar una prueba de una cola**. Utilizar una prueba de una o dos colas depende de la lógica de la hipótesis que se está estudiando. Por lo tanto, al igual que con el nivel de significación, es raro que exista gran posibilidad de elección con respecto a este factor.

6. **Utilizar un procedimiento de prueba de hipótesis más sensible**. Esto es adecuado si es que existen alternativas. En el capítulo 15 analizaremos algunas de las opciones de este tipo. Sin embargo, por lo general el investigador comienza con el método más sensible, razón por la cual no se puede hacer mucho más en este sentido.

La tabla 8-4 resume algunos métodos prácticos para aumentar la potencia de un experimento planificado.

Tabla 8-4.
Resumen de métodos prácticos para aumentar la potencia de un experimento planificado.

Características del estudio	Recursos prácticos para aumentar la potencia	Desventajas
Diferencia predicha entre medias poblacionales ($\mu_1 - \mu_2$)	Aumentar la intensidad del procedimiento experimental.	Puede no ser práctico o puede distorsionar el significado del estudio.
Desvío estándar (σ)	Utilizar una población con menor dispersión.	Puede no haber disponible: disminuye la posibilidad de generalizar.
	Utilizar circunstancias de prueba estandarizadas o controladas, o bien, mediciones más precisas.	No siempre resulta práctico.
Tamaño de muestra (N)	Utilizar un tamaño mayor.	No siempre resulta práctico, puede ser costoso.
Nivel de significación (α)	Utilizar un nivel más indulgente de significación (como 0,1).	Aumenta alfa, la posibilidad de error Tipo I.
Pruebas de una cola versus pruebas de dos colas	Utilizar pruebas de una cola.	Puede no ser apropiado para la lógica del estudio.
Tipo de procedimiento de prueba de hipótesis	Utilizar un procedimiento más sensible.	Puede no existir otro disponible o apropiado.

LA IMPORTANCIA DE LA POTENCIA EN LA EVALUACIÓN DE LOS RESULTADOS DE UN ESTUDIO

Al interpretar los resultados de las investigaciones es muy importante comprender el concepto de potencia estadística, como también, qué factores la afectan.

¿Qué papel cumple la potencia cuando un resultado es significativo?: significación estadística versus significación práctica

Hemos aprendido que un estudio con un mayor tamaño del efecto tiene más posibilidades de resultar significativo. Pero también es posible que un estudio con un tamaño del efecto muy pequeño resulte significativo. Es más probable que esto suceda cuando el estudio tiene un alto nivel de potencia debido a otros factores, especialmente en el caso de una muestra de gran tamaño. Analicemos un estudio en el que entre todos los alumnos que rinden el SAT en determinado año, se selecciona al azar una muestra de aquellos 10.000 cuyo primer nombre comienza con cierta letra. Supongamos que su media en el SAT es de 504, en comparación con una media de 500 ($\sigma = 100$) en el SAT de toda la población. Ese resultado sería significativo al nivel 0,001, pero su tamaño del efecto es un minúsculo 0,04. Es decir, la prueba de significación nos indica que podemos estar bastante seguros de que hay un efecto, que la población de alumnos cuyo primer nombre comienza con esa letra tiene puntuaciones en el SAT más altas que la población general de alumnos. Pero en realidad, el efecto no es muy importante; el tamaño del efecto (o simplemente la diferencia de medias) deja en claro que la diferencia es muy leve. Las distribuciones de las dos poblaciones se superponen tanto que sería de poca utilidad en cualquier caso individual saber con qué letra comienza el primer nombre de una persona.

El mensaje que queremos transmitir con este ejemplo es que existen dos pasos en la evaluación de un estudio. Primero, debemos considerar si el resultado es estadísticamente significativo. Si lo es, esto significa que consideramos que existe un efecto real. Luego debemos analizar si el tamaño del efecto es lo suficientemente grande como para que el resultado sea útil o interesante. Este segundo paso es especialmente importante si el estudio tiene posibles implicancias prácticas. (A veces, tratándose de un estudio que prueba asuntos puramente teóricos, puede ser suficiente simplemente tener la certeza de que existe un efecto en una determinada dirección. Cuando veamos las controversias volveremos a tratar este punto).

Si la muestra era pequeña, podemos suponer que un resultado significativo probablemente también es importante en la práctica. Pero si el tamaño de la muestra es muy grande, debemos tener en cuenta directamente el tamaño del efecto, ya que, en un caso así, es bastante probable que sea demasiado pequeña como para ser útil.

Las implicancias de lo que acabamos de decir constituyen, en parte, una paradoja. La mayoría de las personas supone que cuanto más participantes intervienen en un estudio, más importante será su resultado. En algún sentido, la realidad es justamente al revés. Siendo todos los demás factores los mismos, si un estudio con sólo unos pocos participantes resulta ser significativo, esa significación debe ser el resultado de una gran tamaño del efecto. Un estudio con una gran cantidad de participantes, que resulta estadísticamente significativo, puede tener o no un gran tamaño del efecto.

Es importante observar también que, por lo general, no es una buena idea comparar el nivel de significación de dos estudios para determinar cuál de los dos tiene resultados más importantes. Por ejemplo, un estudio con una cantidad pequeña de participantes, que es significativo al nivel 0,05, bien podría tener un gran tamaño del efecto. Al mismo tiempo, un estudio con una gran cantidad de participantes, que es significativo al nivel 0,001, bien podría tener un tamaño del efecto pequeño.

Sin embargo, el nivel de significación efectivamente nos indica algo. Esto es, cuánta seguridad podemos tener en cuanto a poder rechazar la hipótesis nula, es decir, que existe un efecto distinto de cero. Cuanto menor es el nivel p, mayor es la evidencia de un efecto distinto de cero (Frick, 1997). Sin embargo, definitivamente **no** ocurre que, a menor nivel p, mayor es el efecto. Si dos estudios fueran idénticos en todos los demás factores, un nivel p menor significaría un mayor efecto. Pero si los estudios son diferentes, especialmente si son diferentes en cuanto a tamaño de muestra, la relación del nivel p con el tamaño del efecto es ambigua. Un nivel p pequeño podría ser el resultado de un gran tamaño del efecto, pero bien podría ser también la consecuencia de un gran tamaño de muestra. Por lo tanto, el nivel p indica la fuerza de la evidencia en favor de un efecto distinto de cero. El nivel p no indica el tamaño de ese efecto distinto de cero. (Debido a la importancia de este tema, volveremos a retomarlo más adelante en este mismo capítulo).

¿Qué papel cumple la potencia cuando un resultado no es significativo?

En el capítulo 6 vimos que un resultado no significativo es un resultado no concluyente. Sin embargo, a menudo nos gustaría poder llegar a la conclusión de que existe muy poca o ninguna diferencia entre las poblaciones. ¿Puede ocurrir esto?

Analicemos la relación de la potencia con un resultado no significativo. Supongamos que no obtuvimos un resultado significativo y la potencia del estudio era baja. En ese caso, el estudio es no concluyente. No obtener un resultado significativo puede haber sido la consecuencia de que la hipótesis de investigación fuera falsa, pero también puede haber sido el resultado de que el estudio tuviera una potencia demasiado baja (por ejemplo, por tener muy pocos participantes).

Por el otro lado, supongamos que no obtuvimos un resultado significativo pero la potencia del estudio era alta. En ese caso, parece improbable que la hipótesis de investigación sea verdadera. En estos casos (en los que existe una alta potencia) un resultado no significativo es un argumento bastante fuerte contra la hipótesis de investigación. Esto no significa que todas las versiones de la hipótesis de investigación sean falsas. Por ejemplo, es posible que las poblaciones sean sólo levemente diferentes (y que la potencia haya sido calculada suponiendo una gran diferencia).

En síntesis, el resultado no significativo de un estudio con baja potencia es verdaderamente no concluyente. Sin embargo, el resultado no significativo de un estudio con una potencia alta sugiere que, o bien la hipótesis de investigación es falsa o bien existe un efecto menor del que se predijo al calcular la potencia. También volveremos a tratar este tema más adelante.

La tabla 8-5 resume el papel de la significación y el tamaño de la muestra en la interpretación de resultados experimentales:

Tabla 8-5.
Papel de la significación y del tamaño de la muestra en la interpretación de resultados experimentales.

Resultado estadísticamente significativo	Tamaño de muestra	Conclusión
Si	Pequeño	Resultado importante
Si	Grande	Podría o no tener importancia en la práctica
No	Pequeño	No concluyente
No	Grande	Hipótesis de investigación probablemente falsa

POTENCIA, TAMAÑO DEL EFECTO E INTERVALOS DE CONFIANZA

Dado cualquier tamaño del efecto, a mayor potencia, más estrechos son los intervalos de confianza. La razón es la siguiente: dado cualquier tamaño del efecto, cuanto mayor sea la potencia, más estrecha será la distribución de las medias muestrales; y cuanto más estrecha sea la distribución de las medias muestrales, menor será el intervalo de confianza.

Ahondemos un poco en este tema. Primero analicemos la potencia. La principal influencia en la potencia, además del tamaño del efecto, es el tamaño de la muestra. No importa cuál sea el tamaño del efecto. Cuando la potencia es mayor, el tamaño de la muestra es mayor. Si el tamaño de la muestra es mayor, la distribución de medias es más estrecha, ya que la varianza de la distribución de medias es la varianza de la distribución de observaciones individuales dividida por el tamaño de la muestra.

Analicemos ahora los intervalos de confianza. Veamos cómo construiríamos un intervalo de confianza del 95%. El paso principal es determinar en la distribución de medias muestrales los puntos correspondientes a 1,96 desvíos estándar por debajo, y 1,96 desvíos estándar por encima de la media muestral. En términos de las puntuaciones Z, este intervalo es siempre el mismo. Pero su amplitud en puntuaciones originales depende completamente del desvío estándar de la distribución de medias. Si la distribución de medias es estrecha (con una varianza y un desvío estándar pequeños), el intervalo de confianza es estrecho.

Las implicancias son por demás interesantes. Analicemos un estudio no significativo pero con baja potencia debido al pequeño tamaño de muestra. El intervalo de confianza será muy amplio, incluyendo efectos cero o pequeños al igual que efectos muy grandes. Por lo tanto, los resultados son verdaderamente no concluyentes. Por otro lado, analicemos un estudio no significativo pero con alta potencia debido a un gran tamaño de muestra. El intervalo de confianza será muy estrecho, y todos los valores dentro de él representarán un efecto cero o muy pequeño. En ese caso, tendremos mucha más seguridad de que se sustenta algo semejante a la hipótesis nula.

Existe otro punto que vale la pena mencionar con respecto a la relación de los intervalos de confianza con la potencia y el tamaño del efecto. ¡A veces los investigadores indican los intervalos de confianza en torno a los tamaños del efecto! Así, podríamos encontrar un estudio informando un resultado y agregando algo así como "$d = 0{,}34$, 95%, $IC = 0{,}21$ a 0,47." Esto es particularmente común en un procedimiento de investigación especial que combina tamaños del efecto de muchos estudios diferentes. Ese procedimiento, denominado meta-análisis, es tratado en la próxima sección.

META-ANÁLISIS

El meta-análisis es un desarrollo importante de los últimos años en la estadística, que ha tenido un profundo efecto en la psicología. El meta-análisis es un procedimiento que combina resultados de diferentes estudios, incluso resultados para los cuales se utilizan diferentes métodos de medición, con el fin de sacar conclusiones generales. Al combinar resultados, el tema crucial es la combinación de tamaños del efecto. A modo de ejemplo, un psicólogo especializado en temas sociales podría estar interesado en los efectos causados en los prejuicios por las amistades entre personas de distintas razas, tema sobre el cual se han realizado una gran cantidad de encuestas. El estudio proporcionaría un tamaño del efecto general. También indicaría cómo difieren los tamaños del efecto en los estudios realizados en diferentes países o en cuanto a los prejuicios hacia diferentes grupos étnicos. (Para encontrar un ejemplo de este tipo de meta-análisis *véase* Pettigrew, 1997. Para encontrar otro ejemplo de meta-análisis *véase* el cuadro 8-2).

Cuadro 8-2. Magnitudes de efecto de la relajación y la meditación: un meta-análisis descansado.

En la década de 1980, los resultados de las investigaciones sobre la meditación y la relajación fueron objeto de considerable controversia. Varias revisiones tradicionales del material existente con respecto a estas áreas habían arrojado conclusiones contradictorias en cuanto a si alguna de esas técnicas era beneficiosa y, de ser así, cuáles lo eran. Eppley, Abrams y Shear (1989) decidieron estudiar el tema en forma sistemática, realizando un meta-análisis de los efectos de varias técnicas de relajación sobre la angustia crónica (es decir, no una angustia temporaria sino continua).

Eppley y sus colegas eligieron la angustia crónica para su meta-análisis, porque se trata de un problema preciso relacionado con muchos otros temas de salud mental y, aun así, en sí mismo es un tema que muestra bastante coherencia entre una prueba y otra en las que se utiliza la misma medida, como también entre diferentes medidas.

Los investigadores recopilaron el material publicado siguiendo los procedimientos habituales: no sólo leyeron las revistas científicas especializadas sino también libros y tesis doctorales no publicadas. Uno de los aspectos más complicados del meta-análisis es estar seguro de que uno ha encontrado todas las investigaciones relacionadas con el tema.

Para concluir la investigación, Eppley y sus colegas compararon los tamaños del efecto correspondientes a cada uno de los cuatro métodos principales de meditación y relajación que han sido estudiados en investigaciones sistemáticas. El resultado fue que el tamaño del efecto (*d* de Cohen) promedio correspondiente a los 35 estudios sobre MT (meditación trascendental) fue de 0,70 (lo que indica una diferencia promedio de 0,70 desvíos estándar en la medida de angustia entre aquellos que practicaban este procedimiento de meditación y aquellos que eran parte de los grupos de control). Ese tamaño del efecto era significativamente mayor que el tamaño del efecto de cualquiera de los otros grupos. El tamaño del efecto promedio correspondiente a 44 estudios sobre todos los otros tipos de meditación fue 0,28, el de los 30 estudios sobre relajación progresiva fue de 0,38 y el de los 37 estudios sobre otras formas de relajación fue de 0,40.

Pero en realidad, el meta-análisis recién había comenzado. Había muchas sub-variables de interés. Por ejemplo, al observar las diferentes poblaciones de participantes, las personas que habían sido seleccionadas como altamente angustiadas contribuían en mayor grado al tamaño del efecto, y las poblaciones seleccionadas en la prisión y los individuos más jóvenes aparentemente sacaban más provecho de la MT. No se produjo ningún efecto en el tamaño del efecto como resultado de la capacidad de los instructores, las expectativas de los participantes, el hecho de que los participantes fueran voluntarios o seleccionados al azar para las determinadas condiciones, la predisposición del experimentador (los resultados de la MT en realidad eran más fuertes cuando se eliminaban los datos del investigador que estaba aparentemente en favor de la MT), las distintas medidas de angustia y los diseños de la investigación.

Una clave del alto rendimiento de la MT parecía estar en el hecho de que las técnicas que involucraban concentración producían un efecto significativamente menor, mientras que en la MT, un punto muy especial es la enseñanza de un proceso "espontáneo y sin esfuerzo". Los investigadores creían que la otra diferencia podría estar en los mantras o sonidos, utilizados en la MT que, según se dice, provienen de una muy antigua tradición y, además, son seleccionados para cada alumno por el instructor. Existen investigaciones que indican que los diferentes sonidos sí producen diferentes efectos. Y en este meta-análisis, los métodos de meditación que emplean sonidos del sánscrito seleccionados al azar o palabras inglesas seleccionadas para cada participante no obtuvieron los mismos resultados fuertes.

Cualquiera sea la razón, los autores llegaron a la conclusión de que existen "fundamentos para ser optimistas en cuanto a que, al menos algunos procedimientos de tratamiento actuales, pueden efectivamente reducir la angustia crónica". (p. 973) Por lo tanto si el lector es propenso a preocuparse por pequeñas cosas como un examen de estadística, puede tener en cuenta estos resultados.

Las revisiones de recopilaciones de estudios acerca de un tema en particular, a través del meta-análisis, son una alternativa de las tradicionales publicaciones de revisión "narrativa" del material existente. Esas revisiones tradicionales describen y evalúan cada estudio, y luego intentan sacar alguna conclusión general. La cantidad de revisiones realizadas a través del meta-análisis aumentó significativamente durante la década de 1980 (Myers, 1991), y parece estar transformándose en el método estándar para la revisión de material existente. En los últimos años, incluso se han publicado meta-análisis de meta-análisis. Sin embargo, tal como lo refleja la tabla 8-6, estas publicaciones son especialmente comunes en las áreas más aplicadas de la psicología, y menos comunes en las áreas más teóricas.

Tabla 8-6.
Cantidad de artículos meta-analíticos publicados en varias áreas de la psicología (hasta mediados de 1987).

Sub-disciplina	Frecuencia
Educación	115
Terapia psicológica	100
Psicología industrial/empresarial	44
Psicología social	43
Diferencias sexuales	28
Psicología aplicada a la salud	27
Salud mental	26
Personalidad	16
Psicología experimental	13
Psicología del desarrollo	8

Fuente: Cooper, H. M. & Lemke, K. M. (1991), tab. 1. "Sobre el papel del meta-análisis en la psicología social y de la personalidad." *Boletín de Psicología Social y de la Personalidad [Personality and Social Psychology Bulletin]*, 17, 245–251. Copyright, 1991, por la Society for Personality and Social Psychology, [Sociedad de Psicología Social y de la Personalidad] Inc. Reimpreso con autorización de Sage Publications Inc.

CONTROVERSIAS Y LIMITACIONES: CONTINUACIÓN DE LA CONTROVERSIA ACERCA DE LA SIGNIFICACIÓN ESTADÍSTICA. TAMAÑO DEL EFECTO VERSUS SIGNIFICACIÓN ESTADÍSTICA

En los capítulos 6 y 7, abordamos el tema de la controversia que actualmente se desarrolla sobre el valor de las pruebas de significación, incluso el argumento que sostiene que frecuentemente son mal utilizadas. Dijimos que existían dos modos principales de utilizar inapropiadamente las pruebas de significación que preocupan seriamente a los psicólogos, una de las cuales es que los resultados no significativos son interpretados irreflexivamente como evidencia de que en realidad no existe ningún efecto. En vista del material tratado en este capítulo, podemos comprender con mayor claridad por qué este error es realmente un problema. Los resultados no significativos podrían ser consecuencia de un efecto muy pequeño o inexistente, o bien, simplemente de un bajo nivel de potencia del experimento.

En el capítulo 6, dijimos que pospondríamos la discusión del otro modo de utilizar inapropiadamente las pruebas de significación hasta que hubiéramos tratado determinado material en un capítulo posterior. Ese material era el tamaño del efecto, y ahora estamos en condiciones de analizar ese tema postergado. La utilización inadecuada a la cual nos referimos ocurre cuando un resultado significativo es interpretado irreflexivamente como un resultado "importante"; es decir, se confunde significación con un gran tamaño del efecto.

Hablando en forma general, la significación estadística se refiere a la probabilidad de que pudiéramos haber obtenido nuestro patrón de resultados en forma casual. Como lo explicó Frick (1997), la significación se refiere a la fuerza de la evidencia en favor de un efecto distinto de cero. Si nuestro resultado es significativo a nivel 0,05, tenemos una evidencia bastante buena. Si es significativo a nivel 0,01, es una evidencia aún mejor.

Sin embargo, como hemos visto en este capítulo, un resultado significativo puede no ser importante en el sentido de implicar un gran tamaño del efecto. Por ejemplo, si el tamaño de la muestra fuera grande, un resultado con un pequeño tamaño del efecto podría ser estadísticamente significativo a $p < 0{,}001$. En ese caso tendríamos mucha confianza en que el verdadero efecto es distinto de cero, pero el tamaño de ese efecto distinto de cero aún sería muy pequeño. Llegaríamos a la conclusión de que tenemos un efecto real pero muy leve. De modo similar, si el tamaño de la muestra fuera lo suficientemente pequeño, un resultado con un gran tamaño del efecto podría no ser estadísticamente significativo en absoluto. En ese caso, la mejor estimación en cuanto al tamaño del efecto es que la misma es grande. Pero no tendríamos ninguna certeza ni siquiera de que ese efecto realmente existe; podría ocurrir que el efecto fuera muy pequeño o incluso en dirección opuesta.

Una destacada psicóloga, al escribir sobre este problema (Scarr, 1997), observó que la palabra "significativo" es desafortunada, ya que en lenguaje ordinario significa "importante". De hecho, ella recomendó que se cambiara el nombre por algo así como "confiable". (Ese nuevo nombre también presentaría problemas, dado que el término confiabilidad también tiene un significado especial en estadística). En cualquier caso, no es probable que el nombre cambie a corto plazo. Por lo tanto, es importante que al leer o realizar investigaciones psicológicas recordemos la diferencia entre el uso especial que se da en psicología a la palabra **significación**, y la forma en la que se la utiliza en el lenguaje común.

Tal como observamos en el capítulo 6, la mayoría de los psicólogos no consideran que la utilización inapropiada de las pruebas de significación sea razón suficiente para dejarlas de lado. Sostienen, en cambio, que deberíamos realizar un mayor esfuerzo para evitar tales utilizaciones inapropiadas.

Sin embargo, este no es el fin del problema. Muchos de aquellos que se oponen a las pruebas de significación sostienen que, aun cuando son utilizadas apropiadamente, las pruebas de significación no reflejan el verdadero sentido de las investigaciones. Aseguran que la psicología se refiere fundamentalmente al tamaño del efecto, y no se trata de saber si un resultado es distinto de cero. Ya hemos visto una versión de esta discusión en el capítulo 7, con la sugerencia de que los investigadores utilicen los intervalos de confianza en lugar de las pruebas de significación. La versión completa de esa propuesta (que no analizamos en ese momento) es que en realidad deberíamos informar sobre el tamaño del efecto, con un intervalo de confianza apropiado para ese tamaño del efecto.

Además de los argumentos arriba mencionados, aquellos que proponen el uso del tamaño del efecto sostienen que éste suministra información que puede ser comparada con otros estudios, y utilizada para acumular información de estudios independientes como modo de investigación acerca del progreso en determinado campo. Los tamaños del efecto son componentes cruciales del meta-análisis, y muchos de aquellos que proponen el tamaño del efecto, de hecho, no sólo proponen el meta-análisis sino que lo ven como **la** tendencia del futuro en la psicología.

Existen, sin embargo, argumentos contrarios a favor de las pruebas de significación (y en contra del uso exclusivo del tamaño del efecto). Uno de esos argumentos establece que cuando el tamaño de la muestra es pequeño, aún es posible obtener un gran tamaño del efecto por casualidad. Por lo tanto, si estamos interesados en el resultado de un determinado estudio que utilizó una muestra pequeña, las pruebas de significación nos protegen de tomar los resultados de ese estudio demasiado en serio. De manera similar, existen casos en los que un tamaño del efecto muy pequeño es, de todos modos, importante (*véase* el tratamiento de este tema en el capítulo 3). En una situación de ese tipo, es crucial saber si se puede confiar en que el resultado no es casual. Aun así, muchos de aquellos que sostienen estos argumentos están de acuerdo con que se ha exagerado la importancia de la significación. La mayoría sostiene que la significación debería ser calculada e informada siempre, pero que el tamaño del efecto también debería ser calculado y debería dársele más importancia en la discusión de los resultados.

Existe, además, otra posición que sostiene que en algunas circunstancias los tamaños del efecto son engañosos, por lo cual sólo deberíamos confiar en las pruebas de significación. Chow (1988, 1996), por ejemplo, realiza una diferenciación entre las investigaciones orientadas a la aplicación y aquellas orientadas a la teoría. En la investigación aplicada, los psicólogos están interesados en saber el tamaño real del efecto de un programa determinado o el tamaño de la diferencia real entre dos grupos determinados. En esas circunstancias, Chow está de acuerdo con que el tamaño del efecto es una buena idea. Sin embargo, al realizar investigaciones teóricas, Chow sostiene que la situación es bastante diferente. Es en esas situaciones en las que el tamaño del efecto es irrelevante y hasta engañoso.

Analicemos un experimento acerca del efecto de la familiaridad en el reconocimiento de información. El objetivo de este estudio es analizar la forma básica en que la familiaridad afecta el procesamiento de información. Un estudio podría exponer a diferentes personas a palabras conocidas y no conocidas, y observar cuántas milésimas de segundos les lleva reconocerlas. El tamaño del efecto de tal estudio nos diría muy poco con respecto a la interpretación de los resultados del estudio. La interpretación depende de toda clase de detalles sobre cómo se realizó el estudio, como por ejemplo, qué grado de familiaridad o falta de familiaridad tenían las palabras utilizadas, de qué forma específica fueron presentadas las palabras, y aspectos semejantes. Lo que importa en un estudio de este tipo, según Chow, es que a) la predicción de una diferencia en el reconocimiento de palabras conocidas y no conocidas fuera generada a partir de la teoría, b) que los resultados fueran coherentes con lo predicho (según lo demuestre la significación estadística) y que c) así se sustente la teoría.

La investigación no es fundamentalmente teórica sólo en el campo de la psicología cognitiva. Otros ejemplos de investigación esencialmente teórica incluyen estudios experimentales acerca de motivaciones para la atracción interpersonal, de la medida en que los cambios químicos influyen en los procesos nerviosos, de cómo los niños desarrollan el lenguaje, o de cómo varía la memoria con respecto a hechos emocionales y no emocionales.

De hecho, es probable que el equilibrio actual entre la utilización de pruebas de significación y tamaños del efecto se encuentre simplemente en lo que uno podría esperar de los temas que señala Chow. En las áreas de la psicología aplicada se le está dando una importancia creciente al tamaño del efecto, pero en áreas más teóricas de la psicología es más raro ver menciones explícitas de la magnitud del efecto. Nuestra opinión es que incluso en investigaciones orientadas a la teoría, la pérdida potencial (al colocar el énfasis donde no corresponde) que implica la inclusión del tamaño del efecto, probablemente se vea compensada por, entre otros beneficios, la utilidad para futuros investigadores de poder contar con esa información, lo cual les daría la posibilidad de calcular la potencia al planificar sus propios estudios.

POTENCIA Y TAMAÑO DEL EFECTO SEGÚN SE DESCRIBEN EN PUBLICACIONES CIENTÍFICAS

La potencia es tenida en cuenta principalmente por los psicólogos cuando planifican un estudio. (Por ejemplo, con frecuencia es un tema principal en propuestas para tesis y en pedidos de financiación para gobiernos o instituciones). Como hemos visto, la potencia también representa un papel importante cuando los psicólogos evalúan conclusiones de las publicaciones científicas. De hecho, la función de la potencia aparece a veces en artículos publicados al exponer la interpretación de resultados, especialmente en aquellos no significativos. El que sigue es un ejemplo tomado de una de nuestras propias publicaciones.

En esta investigación (Aron et. al., 1997), pusimos a personas extrañas en pareja y les pedimos que charlaran siguiendo una serie de instrucciones diseñadas para ayudarlos a lograr afinidad. Luego de 45 minutos, cada individuo respondió en forma privada algunas preguntas sobre cuánta afinidad sentía con respecto a su compañero. (Combinamos sus respuestas para formar lo que denominamos "compuesto de afinidad"). En uno de los estudios, probamos si la afinidad sería afectada por a) juntar a los extraños basándonos en la concordancia de sus actitudes o b) hacer creer al participante que lo habíamos juntado con alguien que pensábamos que lo apreciaría.

El resultado en ambos casos –la coincidencia de actitudes y la expectativa de ser apreciado– fue que "no hubo diferencias significativas en el compuesto de afinidad" (p. 567). Después, argumentamos que nuestros resultados sugerían que existía muy poco efecto verdadero de esas variables sobre la afinidad:

> Este estudio presentaba una potencia de aproximadamente un 90% de lograr efectos significativos [...] con respecto a las dos variables manipuladas, si en realidad existía un gran efecto de este tipo ($d = 0,8$). De hecho, la potencia es de aproximadamente un 90% si se trata de encontrar al menos un efecto cuasi significativo ($p < 0,10$) de tamaño mediano ($d = 0,5$). Por lo tanto, parece improbable que hubiéramos podido conseguir estos resultados si de hecho existe más que un efecto pequeño..." (p. 567).

Cada vez es más común (aunque sigue siendo una excepción) que las publicaciones mencionen el tamaño del efecto. Por ejemplo, Caspi et al. (1997) analizaron información de un estudio longitudinal a gran escala de una muestra de niños nacidos alrededor de 1972 en Dunedin, Nueva Zelanda. En una de las partes del estudio, Caspi et. al. compararon a los 94 individuos de su muestra que, a los 21 años, eran dependientes del alcohol (claramente alcohólicos), contra los 863 que no lo eran. Los investigadores compararon estos dos grupos en cuanto a las puntuaciones obtenidas en pruebas de personalidad a los 18 años de edad. Sin embargo, dado que el tamaño de la muestra era tan grande, los investigadores sabían que incluso pequeñas diferencias podrían resultar estadísticamente significativas. Por lo tanto, al describir la planificación de su análisis, observaron:

> Además de probar la hipótesis de que las diferencias entre los grupos son estadísticamente significativas, calculamos los tamaños del efecto (d) de las diferencias obtenidas donde, definiéndolas operativamente, $d = 0,2$ es un tamaño del efecto pequeño, $d = 0,5$ es un tamaño del efecto mediano y $d = 0,8$ es una gran tamaño del efecto (Cohen, 1988). (p. 1055).

El siguiente es un ejemplo tomado de la sección "Respuestas" de la publicación:

> Los jóvenes que a los 21 años de edad eran dependientes del alcohol, a los 18 años tuvieron puntuaciones más bajas en cuanto a Tradicionalismo ($d = 0,49$), Prevención de Daños ($d = 0,44$), Control ($d = 0,64$) y Cercanía Social ($d = 0,40$); y más altos en cuanto a Agresión ($d = 0,86$), Alienación ($d = 0,66$), y Reacción de Estrés ($d = 0,50$).

Es más habitual que se informe sobre el tamaño del efecto en los meta-análisis, en los que se combinan y comparan resultados de diferentes publicaciones. Ya hemos dado varios ejemplos de estos estudios meta-analíticos, incluyendo el del cuadro 8-2. A modo de ejemplo de la forma en que estos estudios realmente describen los resultados en función del tamaño del efecto, analicemos un famoso meta-análisis realizado por Shapiro y Shapiro (1983). Ellos revisaron 143 estudios sobre los efectos de psicoterapias que utilizaban razonablemente métodos de sonido. Entre sus resultados existía una comparación de la efectividad de las terapias en general en diferentes tipos de pacientes (a los que denominaban la "categoría objetivo"). La tabla 8-7 ilustra la cantidad de estudios (N), el porcentaje que representa esa cantidad en relación con todos los estudios revisados, el tamaño del efecto promedio y el desvío estándar de los tamaños del efecto. A partir de esta tabla podemos observar que los mayores beneficios de la psicoterapia se encontraron en los estudios que se concentraban en personas con fobias, y los menores beneficios en estudios que se concentraban en personas con angustia y depresión. Sin embargo, sobre la base de las medidas de Cohen, aún el efecto menor era grande.

Tabla 8-7.
Categorías objetivo y tamaño de efecto.

Categorías objetivo	N	%	Tamaño de efecto	SD
Angustia y depresión	30	7	0,67	0,62
Fobias	76	18	1,28	0,88
Problemas físicos y de hábitos	106	26	1,10	0,85
Problemas sociales y sexuales	76	18	0,95	0,75
Angustias por el desempeño	126	30	0,80	0,71

Fuente: Shapiro, D. A. & Shapiro, D. (1983), tab. 5." Investigación comparativa de resultados de terapias: implicancias metodológicas del meta-análisis". *Revista Científica de Psicología de Asesoramiento y Clínica [Journal of Consulting and Clinical Psychology]*, 51, 42–53. Copyright, 1983, por la Asociación Americana de Psicología [American Psychological Association]. Reimpreso con autorización del autor.

RESUMEN

La potencia estadística de un estudio es la probabilidad de que se obtenga un resultado significativo si la hipótesis de investigación es verdadera.

En la prueba de hipótesis, se dice que se cometió un error Tipo I si el investigador rechaza la hipótesis nula cuando en realidad la hipótesis de investigación es falsa. La probabilidad de un error Tipo I se denomina alfa, siendo alfa igual al nivel de significación. Un error Tipo II ocurre cuando el investigador no rechaza la hipótesis nula, pero en realidad la hipótesis de investigación es verdadera. La probabilidad de un error de Tipo II se denomina beta. La probabilidad de no cometer un error Tipo II (1 – beta) es la potencia de un experimento.

Para calcular la potencia (en el caso de una población conocida y de una sola muestra), primero determinamos el punto de corte acorde con el nivel de significación en puntuaciones originales, sobre la distribución comparativa. La potencia es la probabilidad de obtener una media de al menos esa magnitud en la distribución de la población 1 (la población expuesta al tratamiento experimental). Sobre la base de una media hipotética específica de la población 1 (y si se supone una curva normal con la misma varianza conocida que la población 2), se puede determinar la puntuación Z de ese punto de corte en la distribución comparativa. La probabilidad de exceder esa puntuación Z, la potencia del estudio, puede encontrarse en la tabla de áreas bajo la curva normal.

Existen dos factores principales que afectan la potencia: el tamaño del efecto y el tamaño de la muestra. El tamaño del efecto (*d*) toma en cuenta la diferencia predicha entre medias (cuanto mayor es la diferencia, mayor es la magnitud de efecto) y la varianza de la población (cuanto menor es la varianza poblacional, mayor es la magnitud de efecto). El tamaño del efecto es la diferencia entre las medias poblacionales dividida por el desvío estándar de la población. El tamaño del efecto influye en la potencia, ya que a mayor tamaño del efecto, menor es la superposición entre las distribuciones de medias de la población predicha y el área de rechazo de la distribución de medias correspondiente a la población comparativa. Las reglas del tamaño del efecto de Cohen establecen que un efecto de 0,2 es pequeño, de 0,5 es mediano y de 0,8 es grande. El tamaño del efecto es importante en sí mismo, ya que es un medio estandarizado para evaluar y comparar estudios, el cual no está afectado por el tamaño de la muestra o la escala de medición.

A mayor tamaño de muestra, mayor será la potencia; porque a mayor muestra, menor es la varianza de la distribución de medias, de forma tal que para una determinado tamaño del efecto existe menor superposición entre las distribuciones.

La potencia también es afectada por el nivel de significación (cuanto más extremo, tal como 0,01, menor es la potencia) debido a la utilización de una prueba de una o dos colas (con menor potencia en el caso de utilizarse una prueba de dos colas), y por el tipo de procedimiento de prueba de hipótesis utilizado (en el caso ocasional que pueda elegirse el procedimiento).

Las principales aplicaciones prácticas para aumentar la potencia de un experimento planificado consisten en aumentar el tamaño del efecto y el tamaño de la muestra.

Los resultados significativos de un estudio con alta potencia (como puede ser un estudio con una muestra de gran tamaño) pueden no tener importancia práctica. Los resultados no significativos de un estudio con baja potencia (como puede ser un estudio con una muestra de tamaño pequeño) dejan abierta la posibilidad de que aparezcan resultados significativos si se aumenta la potencia.

Con un determinado tamaño del efecto, los estudios con más potencia (es decir, con muestras mayores) tienen intervalos de confianza más cortos. A veces los tamaños del efecto son informados junto con los intervalos de confianza.

El meta-análisis es un procedimiento reciente para combinar sistemáticamente los efectos de estudios independientes, fundamentalmente sobre la base de los tamaños del efecto.

Los psicólogos disienten con respecto a la importancia relativa de la significación en oposición al tamaño del efecto, y en la interpretación de resultados experimentales. Los psicólogos con orientación teórica parecen otorgar mayor importancia a la significación, mientras que los investigadores de temas de aplicación dan mayor importancia al tamaño del efecto.

Los informes sobre investigaciones a veces incluyen argumentaciones acerca de la potencia, especialmente cuando evalúan resultados no significativos. El tamaño del efecto cada vez aparece con más frecuencia en las publicaciones científicas, y es estándar en aquellas que emplean meta-análisis.

Términos clave

- Alfa (α).
- Beta (β).
- Tamaño del efecto (d).
- Reglas del tamaño del efecto.
- Meta-análisis.
- Tablas de potencia.
- Potencia estadística.
- Error Tipo I.
- Error Tipo II.

Ejercicios

Los ejercicios implican la realización de cálculos (con la ayuda de una calculadora). La mayoría de los problemas estadísticos reales se resuelven por computadora, pero aunque exista la posibilidad de utilizarla, es conveniente realizar estos ejercicios manualmente para incorporar el método de trabajo.

Para adquirir práctica en la utilización de una computadora, para resolver problemas estadísticos, se puede utilizar la sección de Computación de cada capítulo, publicada en la *Guía de estudio y libro de tareas de computación para el alumno [Student's Study Guide and Computer Workbook]* que acompaña este libro.

Todos los datos de esta sección son ficticios (a menos que se especifique lo contrario).

Las respuestas a los ejercicios de la serie I se encuentran al final del libro.

SERIE I

1. Defina alfa y beta.

2. Para cada uno de los siguientes estudios, realice un cuadro de las cuatro posibles decisiones correctas e incorrectas y explique qué significaría cada una de ellas. (Cada cuadro debería estar diseñado de manera semejante a la tabla 8-1, pero dentro de los cuadros debe incluir los resultados reales utilizando los nombres de las variables involucradas en el estudio).

a) Estudio sobre si el aumento del tiempo de descanso mejora el comportamiento de los alumnos en la clase.

b) Estudio sobre si los individuos daltónicos distinguen mejor los matices del gris que la población en su totalidad.

c) Estudio comparativo de los individuos que alguna vez han asistido a psicoterapia y el público en general, para observar si son más tolerantes con las perturbaciones de los demás que la población en general.

3. Aquí le presentamos información acerca de diferentes posibles versiones de un experimento, cada una de las cuales involucra una sola muestra. (Se supone que el investigador puede tener cierto control sobre el desvío estándar y la media predicha de la población cambiando los procedimientos). Determine la potencia y el tamaño del efecto de cada una; luego realice un diagrama de las distribuciones

que se superponen, mostrando las áreas que representan alfa, beta y la potencia. (Suponga que todas las poblaciones tienen una distribución normal).

	Población		Media predicha	N	Nivel de significación	Una o dos colas
	μ	Σ				
(a)	90	4	91	100	0,05	1
(b)	90	4	92	100	0,05	1
(c)	90	2	91	100	0,05	1
(d)	90	4	91	16	0,05	1
(e)	90	4	91	100	0,01	1
(f)	90	4	91	100	0,05	2

4. Basándose en una determinada teoría acerca de la creatividad, un psicólogo predice que los artistas son personas más dispuestas a correr riesgos que la población en general. La población general presenta una distribución normal con una media de 50 y un desvío estándar de 12, según el cuestionario sobre riesgo que el psicólogo piensa utilizar. El psicólogo espera que los artistas tengan un valor promedio de 55, según ese mismo cuestionario. El psicólogo piensa analizar a 36 artistas y probar la hipótesis a un nivel de 0,05. ¿Cuál es la potencia de este estudio? Explique su respuesta a alguien que comprende la prueba de hipótesis con medias muestrales pero que nunca ha aprendido el concepto de potencia.

5. Usted lee un estudio en el que el resultado es apenas significativo a nivel 0,05. Después observa el tamaño de la muestra. Si la muestra es muy grande (en lugar de muy pequeña), ¿cómo debería afectar esto su interpretación de a) la probabilidad de que la hipótesis nula sea realmente verdadera y b) la importancia práctica del resultado? Explique su respuesta a una persona que comprende la prueba de hipótesis pero que nunca ha aprendido el concepto de potencia.

6. ¿Cuál es el efecto "en la potencia de un estudio" de cada uno de los siguientes aspectos?

a) Una mayor diferencia predicha entre las medias poblacionales.

b) Un mayor desvío estándar poblacional.

c) Un mayor tamaño de muestra.

d) Utilizar un nivel de significación más exigente (p. ej. 0,01, en lugar de 0,05).

e) Utilizar una prueba de dos colas en lugar de una.

7. Enumere dos situaciones en las que sea útil tener en cuenta la potencia, indicando cuál es la utilidad de cada una.

SERIE II

1. ¿Qué significa la potencia estadística de un experimento?

2. Para cada uno de los siguientes estudios realice un cuadro de las cuatro posibles decisiones correctas e incorrectas, y explique qué significaría cada una. (Cada cuadro debería estar diseñado de manera semejante a la tabla 8-1, pero dentro de los cuadros debe incluir el resultado real utilizando los nombres de las variables involucradas en el estudio).

a) Estudio sobre si las criaturas que nacen prematuramente comienzan a reconocer los rostros después de lo que lo hacen los demás niños en general.

b) Estudio sobre si los alumnos secundarios que reciben programas de prevención del HIV en sus colegios tienen mayor probabilidad de practicar sexo seguro que otros alumnos secundarios.

c) Estudio sobre si la memoria para ideas abstractas se reduce si la información se presenta en colores que distraen la atención.

3. Aquí le presentamos información sobre diferentes posibles versiones de un experimento planificado, cada una referida a una sola muestra. (Se supone que el investigador puede tener cierto control sobre el desvío estándar y la media predicha de la población cambiando los procedimientos). Determine la potencia y el tamaño del efecto de cada una. Después realice un diagrama de las distribuciones que se superponen mostrando las áreas que representan alfa, beta y la potencia. (Suponga que todas las poblaciones tienen una distribución normal).

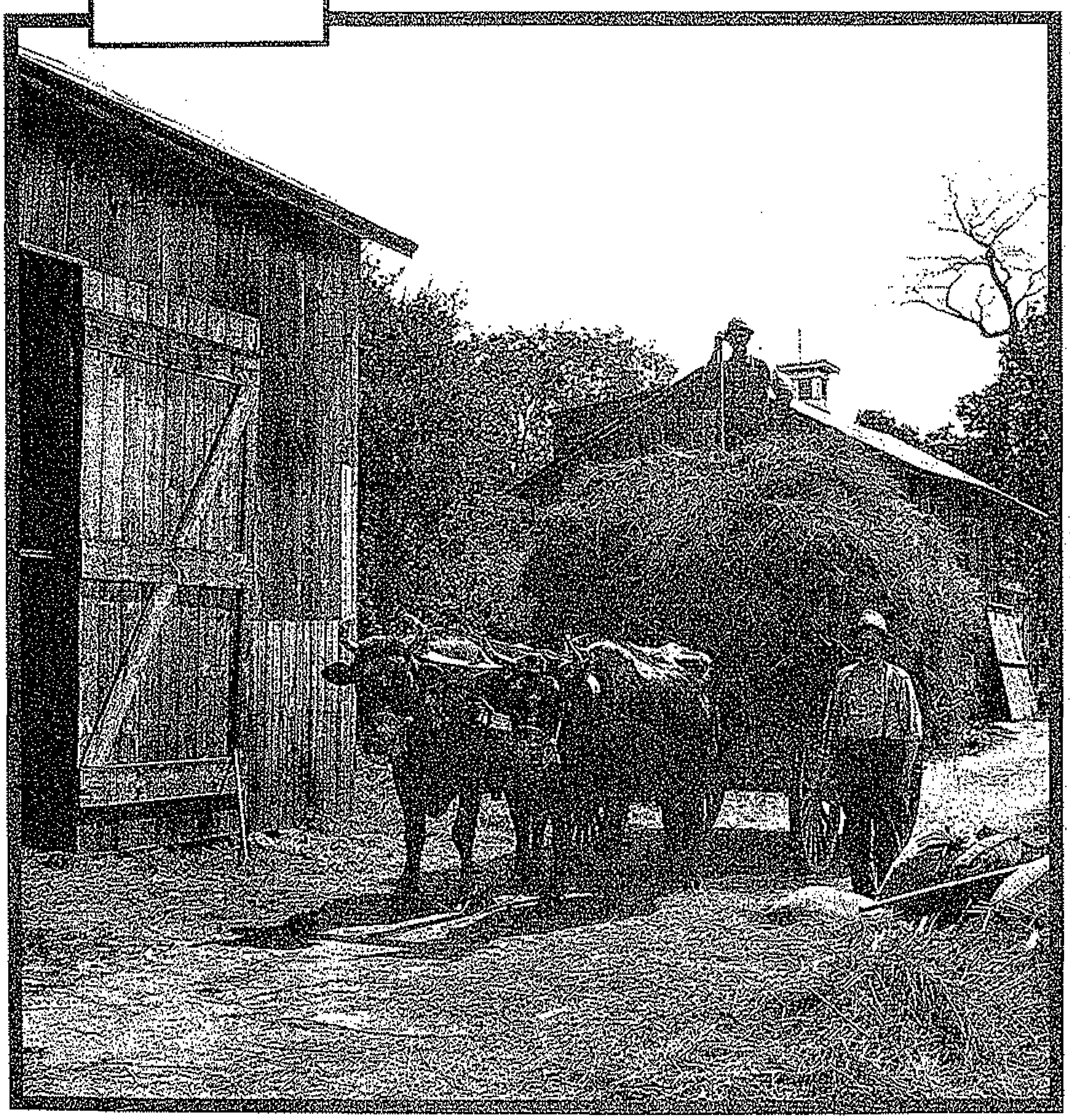
	Población		Media predicha	N	Nivel de significación	Una o dos colas
	μ	Σ				
a)	0	0,5	0,1	50	0,05	1
b)	0	0,5	0,5	50	0,05	1
c)	0	0,5	10,0	50	0,05	1
d)	0	0,5	0,5	100	0,05	1
e)	0	0,5	0,5	200	0,05	1
f)	0	0,5	0,5	400	0,05	2

4. Un psicólogo está planificando un estudio acerca del efecto de la motivación en el desempeño de un participante, en una tarea de atención que involucra la identificación de ciertas letras en una sucesión de letras que pasan a gran velocidad. El investigador sabe, por su larga experiencia, que en condiciones experimentales ordinarias, la población de alumnos que participan en esta tarea identifica, en promedio, 71 de las letras claves (de 100 que se presentan); que el desvío estándar es 10, y que la distribución es aproximadamente normal. El psicólogo predice que si al participante se le paga un dólar por cada letra identificada correctamente, la cantidad media identificada correctamente aumentará a 74. El psicólogo planea probar 20 participantes con estas condiciones, utilizando el nivel 0,05. ¿Cuál es la potencia de este estudio? Explique su respuesta a alguien que comprende la prueba de hipótesis con medias muestrales pero que nunca ha aprendido el concepto de potencia.

5. Usted lee un estudio que, por muy poco, no arroja resultados significativos al nivel 0,05. Es decir, el resultado no es significativo. Después, observa el tamaño de la muestra. Si la muestra es muy grande (y no muy pequeña), ¿cómo afecta esto su interpretación de a) la probabilidad de que la hipótesis nula sea realmente verdadera y de b) la probabilidad de que la hipótesis nula sea realmente falsa? Explique sus respuestas a una persona que comprende la prueba de hipótesis pero que nunca ha aprendido el concepto de potencia.

6. Usted está planificando un estudio que, de acuerdo con sus cálculos, tiene una potencia bastante baja. Nombre seis alternativas de las que dispone para aumentar la potencia.

9 Prueba *t* para medias dependientes



Descripción del capítulo

A esta altura, el alumno debe creer que lo sabe todo acerca de la prueba de hipótesis. Sin embargo, se sorprenderá: lo que ha aprendido hasta ahora no le resultará muy útil como psicólogo. ¿Por qué? Los procedimientos para prueba de hipótesis descriptos hasta ahora fueron, por supuesto, requisitos previos absolutamente necesarios para lo que estamos por aprender. Sin embargo, estos procedimientos involucraban la comparación de un grupo de valores observados con una población conocida, y cuando se realizan investigaciones reales, con frecuencia se comparan dos o más grupos de valores observados entre sí, sin ninguna información directa acerca de las poblaciones. Por ejemplo, podrían utilizarse dos valores correspondientes a cada una de las diferentes personas, tales como las puntuaciones en una prueba de angustia antes y después de la psicoterapia; o la cantidad de palabras familiares recordadas, en comparación con las no familiares, en un experimento acerca de la memoria. O también se podría utilizar un valor por cada una de las personas que forman dos grupos, tales como un grupo experimental y un grupo control, en un estudio acerca del efecto de la pérdida del sueño en la resolución de problemas.

Estos tipos de situaciones de investigación se encuentran entre las más comunes en psicología, donde la única información disponible proviene de las muestras. Nada se sabe acerca de las poblaciones de donde provienen esas muestras. Particularmente, el investigador desconoce la varianza de las poblaciones involucradas, la cual es un componente crucial en el paso 2 del proceso de prueba de hipótesis (determinar las características de la distribución comparativa).

En este capítulo, analizamos primero la solución al problema de no conocer la varianza poblacional. Comenzamos con una situación de prueba de hipótesis especial, comparando la media de una sola muestra con una población de la cual conocemos la media pero no la varianza. Luego, después de haber aprendido cómo se maneja este inconveniente de no conocer la varianza poblacional, proseguimos con la situación en la cual directamente no hay población conocida, una situación en la que todo lo que tenemos son dos observaciones por cada una de las personas de un grupo.

Los procedimientos de prueba de hipótesis que aprenderemos en este capítulo, en los que no se conoce la varianza poblacional, son ejemplos de lo que se denominan pruebas *t*. **Las pruebas *t*** a veces se denominan "*t* de *Student*", porque sus principios fundamentales fueron desarrollados originalmente por William S. Gosset, quien publicó sus artículos bajo el seudónimo de "*Student*" (*véase* cuadro 9-1).

Cuadro 9-1. William S. Gosset, alias "*Student*": no era un matemático sino un "hombre práctico".

William S. Gosset se graduó en Oxford en 1899 y obtuvo su diploma en matemática y química. En el mismo año sucedió que los fabricantes de cerveza de Guinness, en Dublín, Irlanda, estaban buscando científicos jóvenes para que, por primera vez en la historia, analizaran la fabricación de la cerveza de manera científica. Gosset obtuvo uno de esos puestos y no tardó en sumergirse en la cebada, los lúpulos y cubas para la elaboración de la cerveza.

El problema consistía en hallar la forma de que la calidad de la cerveza fuera menos variable, y especialmente descubrir la causa de las malas tandas. Un científico que se preciara de serlo recomendaría, sin duda, la realización de experimentos. Pero un negocio como el de la elaboración de cerveza no podía darse el lujo de gastar dinero en experimentos que incluían grandes cantidades de cubas, algunas de las cuales iban a perderse, como lo sabría cualquier fabricante de cerveza. Por lo tanto, Gosset se vio forzado a analizar la probabilidad de qué cierta especie de cebada produjera una cerveza de pésima calidad, dado que el experimento podía consistir sólo en unas pocas tandas de cada especie. A este problema se sumaba el hecho de que él no tenía la menor idea de la variabilidad de las especies de cebada: tal vez algunos campos dieran mejor cebada al ser plantadas con la misma especie (¿suena familiar?). Pobre Gosset, al igual que los psicólogos de nuestro tiempo, no tenía idea de la varianza de su población.

Gosset estaba a la altura de las circunstancias, aunque en ese momento sólo él lo sabía. Para sus colegas de la fábrica de cerveza, era un profesor de matemática y no un digno fabricante de cerveza. Para sus colegas estadísticos, principalmente los del Laboratorio de Estadística de Datos Biológicos de la Universidad de Londres, era un simple fabricante de cerveza y no un matemático propiamente dicho. En resumen, Gosset era esa clase de científico que no tiene inconveniente en aplicar sus talentos a la vida práctica.

De hecho, parecía disfrutar de esa vida real: cultivando peras, pescando, jugando golf, construyendo botes, esquiando, andando en bicicleta (y jugando a las bochas sobre césped, después de que se quebró la pierna al estrellar su auto –un Ford modelo T de dos plazas al que llamaba "la cama voladora"– contra un poste de luz). Disfrutaba especialmente de las herramientas simples que podían aplicarse a cualquier cosa, fórmulas simples que podía calcular mentalmente. (Un amigo lo describía como un experto carpintero, aunque afirmaba que Gosset realizaba casi toda su carpintería fina sólo con un cortaplumas).

De esa manera, Gosset descubrió la distribución *t* e inventó la prueba *t* (la simpleza misma, comparada con la mayoría de los cálculos estadísticos), para aquellas situaciones en las que las muestras son pequeñas y se desconoce la variabilidad de la población que se supone de un tamaño mucho más grande. La mayor parte de su trabajo lo realizó en el reverso de sobres, con muchos errores menores de aritmética que tuvo que corregir luego. Como suele ocurrir, publicó su trabajo sobre "Métodos para la elaboración de cerveza" sólo cuando algunos editores de las revistas científicas se lo pidieron. Hasta el día de hoy, la mayoría de los estadísticos llaman a la distribución *t* la "*t* de *Student*", porque Gosset escribía bajo el seudónimo de "*Student*", simplemente para que la fábrica de cerveza Guinness no tuviera que admitir públicamente que a veces elaboraban una mala tanda de cerveza.

Referencias: Peters (1987); Stigler (1986); Tankard (1984).

INTRODUCCIÓN A LA PRUEBA *T*: PRUEBA *T* PARA UNA SOLA MUESTRA

Comenzaremos con la siguiente situación: tenemos los registros de una sola muestra y queremos comparar esos datos con una población de la cual conocemos la media pero no la varianza. La prueba de hipótesis, en este caso, se denomina **prueba *t* para una sola muestra**. (También la llaman "prueba *t* de una muestra"). La prueba *t* para una sola muestra funciona básicamente de la misma forma que lo aprendido en el capítulo 7. Hay sólo dos importantes cuestiones nuevas: primero, ya que no conocemos la varianza poblacional, debemos estimarla. Segundo, cuando se debe estimar la varianza de la población, la forma de la distribución comparativa es levemente diferente a una curva normal.

Ejemplo

Supongamos que el periódico de cierta facultad informa acerca de una encuesta informal que muestra que los estudiantes de la facultad estudian un promedio de 2,5 horas por día. Sin embargo, uno de los alumnos considera que los estudiantes que viven en el mismo alojamiento estudiantil que él estudian mucho más que esa cantidad de horas. Elige al azar 16 alumnos del edificio y les pregunta cuánto estudian cada día. (Supondremos que son todos honestos y precisos). El resultado que obtiene es que estos 16 alumnos estudian un promedio de 3,2 horas por día. En ese caso, ¿el alumno debería concluir que los estudiantes de su alojamiento estudian más que el promedio de horas que lo hacen los de la facultad? ¿O debería concluir que sus resultados son tan cercanos a ese promedio de la facultad que la pequeña diferencia de 0,7 horas podría bien deberse a que accidentalmente ha seleccionado 16 de los residentes más estudiosos del alojamiento estudiantil?

El primer paso del proceso de prueba de hipótesis es replantear el problema en función de hipótesis sobre poblaciones. Existen dos poblaciones:

Población 1: el tipo de estudiantes que viven en el edificio del alumno que realiza el estudio.
Población 2: el tipo de estudiantes de la facultad en general.

La hipótesis de investigación establece que los alumnos de la población 1 estudian más que los alumnos de la población 2; la hipótesis nula establece que los alumnos de la población 1 no estudian más que los alumnos de la población 2. Hasta aquí el problema no es diferente al del capítulo 7.

El segundo paso es determinar las características de la distribución comparativa. La media de esta distribución será de 2,5, el número arrojado por la encuesta a los alumnos de la facultad en general (población 2).

La siguiente parte del segundo paso es encontrar la varianza de la distribución de medias. En este ejemplo nos encontramos con otro tipo de inconveniente; hasta aquí, siempre hemos conocido la varianza de la población de observaciones individuales. Utilizando esa varianza, luego calculábamos la varianza de la distribución de medias. En este caso, la publicación no informó la varianza de la cantidad de horas de estudio de la facultad en general. Entonces el alumno llama al periódico. Lamentablemente, el periodista no calculó la varianza, y los resultados de la encuesta original ya no están disponibles. ¿Qué hacer en ese caso?

Principio básico de la prueba *t*: estimar la varianza poblacional a partir de los valores muestrales

Si no conocemos la varianza de la población de observaciones, la podemos estimar a partir de lo que sí conocemos: los valores observados de las personas que forman la muestra. Según la lógica de la prueba de hipótesis, se considera que el grupo de personas que analizamos es una muestra aleatoria de determinada población. La varianza de esa muestra debería reflejar la varianza de la población. Si la población presenta mucha dispersión (existe mucha varianza entre los valores), entonces una muestra seleccionada al azar de esa población debería tener mucha dispersión; si la población es muy compacta, con poca dispersión, no debería haber mucha dispersión tampoco en la muestra. Por lo tanto, se podría utilizar la dispersión de los valores de la muestra para realizar una presunción fundamentada de la dispersión de los valores de la población. Es decir, podríamos calcular la varianza de las observaciones muestrales, y ese cálculo sería similar a la varianza de los valores poblacionales. (*Véase* figura 9-1).

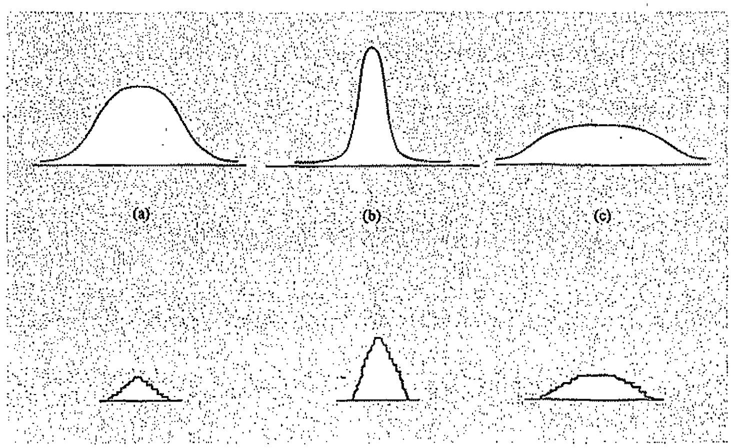


Figura 9-1. Varianzas muestrales y las poblaciones de donde provienen las muestras.

Sin embargo, existe un pequeño obstáculo. La varianza de una muestra generalmente será levemente menor que la varianza de la población de donde proviene la muestra. Por está razón, la varianza de la muestra es una **estimación sesgada** de la varianza poblacional.

¿Por qué la varianza de la muestra es levemente menor que la de la población? La varianza se basa en desvíos con respecto a la media. La varianza de una población se basa en desvíos con respecto a la media de esa población. Por otro lado, la varianza de una muestra se basa en desvíos con respecto a la media de esa muestra. La media de una muestra es el punto de equilibrio óptimo para sus registros. Por lo tanto, los desvíos de los registros de una muestra con respecto a su media serán menores que los desvíos con respecto a cualquier otro número. La media de la muestra generalmente no es exactamente igual a la media de la población de donde proviene. Consecuentemente, los desvíos de los registros de una muestra con respecto a la media de la muestra generalmente serán menores que los desvíos de los registros de esa muestra con respecto a la media de la población.

Supongamos que conociéramos la media poblacional de la que proviene la muestra y utilizáramos esta media para comparar el desvío de cada registro de la muestra. La varianza calculada de este modo sería una **estimación no sesgada de la varianza poblacional**.

Lamentablemente, no conocemos la media de la población de la cual proviene la muestra. La muestra proviene de la población 1. En este caso, sólo conocemos la media de la población 2. Pero las medias de las dos poblaciones son iguales sólo si la hipótesis nula es verdadera, y eso es precisamente lo que estamos probando. (Independientemente de si la hipótesis nula es verdadera o no, nosotros **sí** suponemos que ambas poblaciones tienen la misma varianza).

Afortunadamente, podemos calcular una estimación no sesgada de la varianza poblacional. Lo que hacemos es realizar una corrección al calcular la varianza, basadas en los valores muestrales que refleja con exactitud la medida en que la media de una muestra tiende a variar con respecto a la verdadera media de la población. Esta estimación no sesgada se calcula cambiando ligeramente la fórmula ordinaria de varianza. La manera común de calcular la varianza es tomar la suma de los desvíos cuadráticos y dividirla por la cantidad de valores observados. Según el procedimiento modificado, tomamos la suma de los desvíos cuadráticos pero la dividimos por la cantidad de valores **menos 1.** Dividir por una cantidad ligeramente menor hace que el resultado de la división (la varianza) sea ligeramente mayor.

Sucede que dividir por la cantidad de valores menos 1 aumenta la varianza resultante sólo lo suficiente como para que sea una estimación no sesgada de la varianza poblacional. A propósito, "no sesgada" no significa que la estimación será exactamente la verdadera varianza de la población; sólo significa que el método produce estimaciones cuyo promedio coincide con esa verdadera varianza. (La estimación sesgada, la varianza muestral calculada en la forma usual, será sistemáticamente demasiado baja).

El símbolo de la estimación no sesgada de la varianza poblacional es S^2. La fórmula es la usual, pero con la división por $N-1$ en lugar de N:

$$S^2 = \frac{\Sigma(X-M)^2}{N-1} = \frac{SC}{N-1} \tag{9-1}$$

el desvío estándar poblacional estimado es la raíz cuadrada de la varianza poblacional estimada,

$$S = \sqrt{S^2} \tag{9-2}$$

Volvamos al ejemplo de las horas de estudio y calculemos la varianza poblacional estimada utilizando los 16 valores muestrales. Primero, calculamos la suma de los desvíos cuadráticos. (Restamos la media muestral a cada uno de los valores, elevamos al cuadrado esos desvíos, y los sumamos). Supongamos que realizamos este cálculo y el resultado es 9,6 ($SC = 9{,}6$). Para obtener la varianza poblacional estimada, dividimos esta suma de desvíos cuadráticos por la cantidad de valores muestrales menos 1. En la muestra hay 16 valores, entonces el tamaño de la muestra menos 1 es 15. El resultado es 0,64. Es decir, 9,6/15 es igual a 0,64. La fórmula es la siguiente:

$$S^2 = \frac{\Sigma(X-M)^2}{N-1} = \frac{SC}{N-1} = \frac{9{,}6}{16-1} = \frac{9{,}6}{15} = 0{,}64$$

Grados de libertad

El mínimo por el cual dividimos (la cantidad de valores menos 1) para calcular la varianza poblacional estimada tiene un nombre especial. Se lo denomina **grados de libertad**, porque es la cantidad de valores muestrales "libres para variar". Se trata de un concepto un poco complicado. La idea básica es que, al calcular la varianza, primero debemos conocer la media; si conocemos la media y todos los valores de la muestra excepto uno, con un poco de aritmética podemos calcular aquél valor que desconocemos. (Si al alumno le agradan las aventuras matemáticas, puede intentarlo con algunos ejemplos para comprobar como funciona). Por lo tanto, una vez que conocemos la media, uno de los valores de la muestra no tiene libertad de tomar cualquier valor posible. Entonces, los grados de libertad son la cantidad de valores menos 1. Se expresa por la fórmula,

$$gl = N - 1 \tag{9-3}$$

donde gl representa los grados de libertad. En nuestro ejemplo, $gl = 16 - 1 = 15$. (En algunos casos, que aprenderemos en capítulos posteriores, los grados de libertad se calculan de forma ligeramente diferente, debido a que en esos casos es diferente la cantidad de valores libres para variar. En todos los casos planteados en este capítulo, $gl = N - 1$).

La fórmula para calcular la varianza poblacional estimada, con frecuencia, se escribe utilizando gl en lugar de $N - 1$:

$$S^2 = \frac{\Sigma(X-M)^2}{gl} = \frac{SC}{gl} \tag{9-4}$$

Determinación del desvío estándar de la distribución de medias a partir de una varianza poblacional estimada

Una vez que hemos estimado la varianza de la población, calcular el desvío estándar de la distribución comparativa implica los mismos procedimientos aprendidos en el capítulo 7. Es decir, consideramos la distribución comparativa como una distribución de medias. Al igual que antes, podemos calcular su varianza como la varianza de la población de individuos dividida por el tamaño de la muestra. La única diferencia es que en lugar de conocer la varianza de la población de observaciones individuales hemos tenido que estimarla. Como siempre, el desvío estándar de la distribución de medias es la raíz cuadrada de su varianza. La fórmula es la siguiente,

$$S_M^2 = \frac{S^2}{N} \tag{9-5}$$

$$S_M = \sqrt{S_M^2} \tag{9-6}$$

Es importante tener en cuenta que cuando estamos utilizando una varianza poblacional estimada, los símbolos para la varianza y el desvío estándar de la distribución de medias utilizan S en lugar de σ.

En el ejemplo que estamos analizando, el tamaño de la muestra era 16, y la varianza poblacional estimada que acabamos de calcular era 0,64. La varianza de la distribución de medias, sobre la base de esa estimación, será 0,04. Es decir, 64 dividido 16 es igual a 0,04. El desvío estándar es 0,2, la raíz cuadrada de 0,04. La fórmula es la siguiente,

$$S_M^2 = \frac{S^2}{N} = \frac{0,64}{16} = 0,04$$

$$S_M = \sqrt{S_M^2} = \sqrt{0,04} = 0,2$$

Cabe advertir que para encontrar la varianza de una distribución de medias siempre se divide la varianza poblacional por el tamaño de la muestra, y esto ocurre ya sea porque conocemos la varianza de la población o sólo porque la estimemos. En el ejemplo que estamos analizando, dividimos la varianza poblacional, que habíamos estimado, por 16. Sólo cuando realizamos la estimación de la varianza poblacional dividimos por el tamaño de la muestra menos 1. Es decir, los grados de libertad se utilizan sólo cuando estimamos la varianza de la población de observaciones individuales.

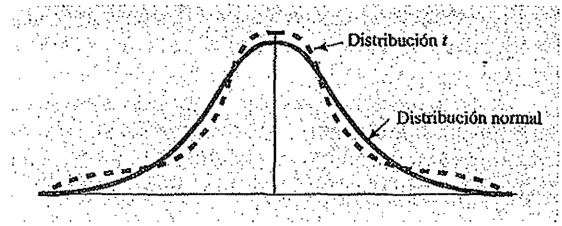
Forma de la *distribución comparativa* al *utilizar una varianza* **poblacional estimada:** la distribución *t*

En el capítulo 7 dijimos que mientras sea razonable suponer que la distribución poblacional sigue una curva normal, la forma de la distribución de medias también seguirá una curva normal. Esto cambia cuando estamos realizando una prueba de hipótesis utilizando una varianza poblacional estimada. Cuando eso ocurre, contamos con menos información cierta y existe más posibilidad de error. El efecto matemático es que las medias extremas son ligeramente más probables que en una curva normal. Más aún, cuanto menor sea el tamaño de la muestra, mayor será esa tendencia, ya que estamos estimando la varianza de la población basándonos en menos información.

¿Cuál es el resultado de todo lo anterior cuando realizamos una prueba de hipótesis con una varianza estimada? El resultado es que la distribución de medias (la distribución comparativa) no seguirá exactamente una curva normal. Por el contrario, la distribución comparativa sigue una curva matemáticamente definida que se denomina **distribución *t*.**

En realidad, existen muchas distribuciones *t*. Su forma varía según los grados de libertad de la muestra utilizada al estimar la varianza poblacional. (Sin embargo, para un determinado grado de libertad, existe sólo una distribución *t*). En general, todas las distribuciones *t* parecen a la vista una curva normal, con forma de campana, completamente simétricas y unimodales. La distribución *t* se diferencia ligeramente porque sus colas son más gruesas (es decir, existen algunos valores más en los extremos). La figura 9-2 ilustra la forma de una distribución *t* en comparación con una curva normal.

Figura 9-2. Distribución *t* comparada con la distribución normal.



Esta sutil diferencia de la forma afecta los valores extremos necesarios para rechazar la hipótesis nula. Para rechazar la hipótesis nula necesitamos estar en una zona extrema bajo la curva normal, como por ejemplo el 5% superior. Sin embargo, si hay más valores extremos, el punto en el que comienza el 5% superior está más alejado, hacia afuera de la curva. Por eso, es necesaria una media *muestral más extrema para obtener significación al utilizar una distribución t que al utilizar una curva normal.*

La medida en que la distribución *t* difiere de la curva normal depende precisamente de los grados de libertad en la estimación de la varianza poblacional. La distribución *t* difiere más de la curva normal cuando la estimación de la varianza poblacional se basa en una muestra muy pequeña, de modo que los grados de libertad son bajos. Por ejemplo, utilizando la curva normal, el punto de corte para una prueba de una cola a nivel 0,05 es 1,64. En una distribución *t* con 7 grados de libertad (es decir, con un tamaño de muestra de 8), el punto de corte correspondiente al 5% en una prueba de una cola es 1,895. Si la varianza poblacional estimada se basa en una muestra mayor, digamos una muestra de 25 (de modo que $gl = 24$), el punto de corte es 1,711. Si el tamaño de la muestra es infinito, la distribución *t* es igual a la curva normal. (Por su puesto, si el tamaño de tu muestra fuera infinito, ¡incluiría toda la población!). Pero incluso con tamaños de muestra de *30 ó más, la distribución t es casi* idéntica a la curva normal.

Antes de aprender cómo encontrar realmente el punto de corte utilizando una distribución *t*, volvamos primero brevemente al ejemplo de la cantidad de horas que estudian cada noche los alumnos del edificio de dormitorios . Finalmente tenemos todo lo que necesitamos para completar el segundo paso, que se refiere a las características de la distribución comparativa. Ya hemos visto que la distribución de medias tendrá una media de 2,5 horas y un desvío estándar de 0,2. Sobre la base de lo que acabamos de analizar, ahora podemos agregar que la forma de la distribución comparativa será una distribución *t* con 15 grados de libertad.[1]

[1] Los estadísticos hacen una sutil distinción en este caso entre la distribución comparativa y la distribución de medias. Hemos evitado presentar esta distinción aquí y en capítulos posteriores para simplificar el tratamiento de un tema que ya es de por sí bastante complicado. Pero para aquellos que estén interesados en el tema diremos que la distinción puede entenderse de la siguiente manera: el procedimiento general de prueba de hipótesis, tal como lo presentamos en el capítulo 7, puede describirse como la comparación de una puntuación *Z* con la media de la muestra, donde $Z = (M - \mu)/\sigma_M$ y donde $\sigma_M = \sqrt{\sigma^2/N}$, y luego la comparación de esta puntuación *Z* con un punto *Z* de corte de la tabla de áreas de la curva normal. Describimos este proceso utilizando la distribución de medias como distribución comparativa.
Los estadísticos dirían que en realidad estamos comparando la puntuación Z, calculada con una distribución de la puntuación Z (que es simplemente una curva normal estándar). De modo similar, en el caso de una prueba *t*, los estadísticos consideran que el procedimiento es como calcular una puntuación *t* (similar a una puntuación Z pero calculada utilizando un desvío estándar estimado), donde $t = (M - \mu)/S_M$, donde $S_M = \sqrt{S^2/N}$, y luego comparar la puntuación *t* calculada con un punto de corte *t* tomado de una tabla de distribución *t*. Por lo tanto, de acuerdo con la lógica estadística formal, la distribución comparativa es una distribución de la puntuación *t*, y no de medias.

Determinación del valor muestral de corte para rechazar la hipótesis nula: utilización de la tabla *t*

El tercer paso del proceso de prueba de hipótesis es determinar el punto de corte para rechazar la hipótesis nula. Existe una distribución *t* diferente para cada número de grados de libertad en particular. Sin embargo, para no llenar hojas y hojas con tablas para cada posible distribución *t*, se utiliza una tabla simplificada que incluye sólo los puntos de corte cruciales. En el apéndice B incluimos esta **tabla *t*** (tabla B-2).

En el ejemplo que estamos analizando, tenemos una prueba de una cola (nos interesa saber si los alumnos del edificio en cuestión estudian **más** que los alumnos de esa facultad en general). Creemos que podemos utilizar el nivel de significación del 5% ya que el costo de un error Tipo I (rechazar equivocadamente la hipótesis nula) no es grande. Tenemos 16 participantes, lo que da 15 grados de libertad para la estimación de la varianza poblacional.

La tabla 9-1 incluye una parte de una tabla *t* similar a la tabla B-2. Buscamos la columna correspondiente al nivel de significación 0,05 para pruebas de una cola, luego descendemos por esa columna hasta la línea correspondiente a 15 grados de libertad. El número crucial de corte es 1,753. Esto significa que rechazaremos la hipótesis nula si la media muestral se encuentra a 1,753 o más desvíos estándar por encima de la media en la distribución comparativa. (Si estuviéramos utilizando una varianza conocida hubiéramos encontrado el punto de corte en una tabla de áreas bajo la curva normal. La puntuación Z necesaria para rechazar la hipótesis nula, sobre la base de la curva normal, hubiera sido 1,645).

Hay otro tema que queremos destacar acerca de la utilización de la tabla *t*. En la tabla *t* completa que se encuentra en el apéndice, existe una línea para cada grado de libertad, desde el 1 hasta el 30. Luego, para cada cinco grados de libertad (35, 40, 45, etc.), hasta 100. Supongamos que el estudio incluyera grados de libertad que se encuentran entre dos valores. Para mayor seguridad, deberíamos utilizar los grados de libertad **inferiores** más cercanos a los de la muestra que se encuentren en la tabla. Por ejemplo, si estuvieras realizando un estudio en el que hubieran 43 grados de libertad, utilizaríamos la línea de la tabla correspondiente a 40 *gl*.

Tabla 9-1.
Puntos de corte para las distribuciones *t* con grados de libertad del 1 a 17. (Se indica el punto de corte para el ejemplo acerca de las horas de estudio).

	Pruebas de una cola			Pruebas de dos colas		
gl	*0,10*	*0,05*	*0,01*	*0,10*	*0,05*	*0,01*
1	3,078	6,314	31,821	6,314	12,706	63,657
2	1,886	2,920	6,965	2,920	4,303	9,925
3	1,638	2,353	4,541	2,353	3,182	5,841
4	1,533	2,132	3,747	2,132	2,776	4,604
5	1,476	2,015	3,365	2,015	2,571	4,032
6	1,440	1,943	3,143	1,943	2,447	3,708
7	1,415	1,895	2,998	1,895	2,365	3,500
8	1,397	1,860	2,897	1,860	2,306	3,356
9	1,383	1,833	2,822	1,833	2,262	3,250
10	1,372	1,813	2,764	1,813	2,228	3,170
11	1,364	1,796	2,718	1,796	2,201	3,106
12	1,356	1,783	2,681	1,783	2,179	3,055
13	1,350	1,771	2,651	1,771	2,161	3,013
14	1,345	1,762	2,625	1,762	2,145	2,977
15	1,341	1,753	2,603	1,753	2,132	2,947
16	1,337	1,746	2,584	1,746	2,120	2,921
17	1,334	1,740	2,567	1,740	2,110	2,898

Determinación del valor correspondiente a la media muestral en la distribución comparativa: el punto *t*

El cuarto paso del proceso de prueba de hipótesis es la determinación del valor muestral en la distribución comparativa. En capítulos anteriores, esto implicaba ubicar la puntuación Z en la distribución comparativa (la cantidad de desvíos estándar a los que se encontraba el valor muestral con respecto a la media en la distribución de medias). Cuando la distribución comparativa es una distribución *t* hacemos exactamente lo mismo. La única diferencia es que antes, cuando la distribución comparativa era una curva normal, el valor que calculábamos en ella se llamaba puntuación Z; ahora estamos utilizando una distribución *t* como distribución comparativa, por lo tanto, el valor que calculamos sobre ella se denomina **puntuación *t*.** La fórmula es la siguiente,

$$t = \frac{M - \mu}{S_M} \tag{9-7}$$

En el ejemplo que estamos analizando, la media muestral de 3,2 está a 0,7 horas de la media de la distribución de medias. Es decir, a un total de 3,5 desvíos estándar de la media (es decir, 0,7 horas dividido por el desvío estándar de 0,2 horas es igual a 3,5). En otras palabras, la **puntuación *t*** en el ejemplo es 3,5. Aplicando la fórmula se obtiene:

$$t = \frac{M - \mu}{S_M} = \frac{3,2 - 2,5}{0,2} = \frac{0,7}{0,2} = 3,5$$

Determinación de rechazar o no la hipótesis nula

El quinto paso de la prueba de hipótesis es comparar los valores de los pasos 3 y 4 para decidir si se rechaza la hipótesis nula. Este paso es exactamente igual en el caso de la prueba *t* que en el caso estudiado en los capítulos anteriores. Comparamos el punto de corte del paso 3 con el punto muestral en la distribución comparativa del paso 4. En el ejemplo que analizamos, el punto *t* de corte era 1,753, y el punto *t* real de nuestra muestra era 3,5. Conclusión: se rechaza la hipótesis nula y se sostiene la hipótesis de investigación que establecía que los alumnos del edificio de dormitorios en cuestión estudian más que los alumnos del resto de la facultad.

La figura 9-3 representa gráficamente las distribuciones del ejemplo que acabamos de analizar.

Resumen de la prueba de hipótesis cuando se desconoce la varianza poblacional

La prueba de hipótesis, cuando se desconoce la varianza de la población, es exactamente igual a la prueba de hipótesis descripta en el capítulo 7, pero con cuatro excepciones: a) en lugar de saber por adelantado la varianza de la población, ésta se estima a partir de la muestra (utilizando la fórmula para la estimación no sesgada, $S^2 = SC/gl$); b) en lugar de seguir una curva normal, la distribución comparativa sigue una distribución *t* con *gl* igual al número de valores observados de la muestra menos 1; c) en lugar de buscar el punto de corte, correspondiente al nivel de significación elegido en una tabla de áreas bajo la curva normal, utilizamos una tabla *t*, y d) el valor muestral en la distribución comparativa, en lugar de llamarse puntuación Z, se denomina puntuación *t*. La tabla 9-2 compara las dos situaciones sistemáticamente.

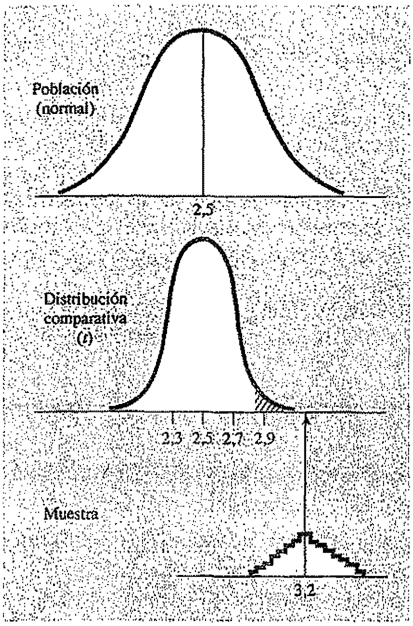


Figura 9-3. Distribuciones relacionadas con el ejemplo de las horas de estudio.

Tabla 9-2.
Prueba de hipótesis con una sola media muestral, y en la que se desconoce la varianza de la población (prueba *t*) en comparación con los casos en los que se conoce la varianza poblacional.

Pasos de la prueba de hipótesis	Diferencia con los casos en los que se conoce la varianza poblacional
1. Replantear el problema en función de hipótesis de investigación e hipótesis nula sobre las poblaciones.	No hay diferencia en el método.
2. Determinar las características de la *distribución comparativa* :	
Media poblacional	No hay diferencia en el método.
Varianza poblacional	Se estima a partir de la muestra.
Desvío estándar de la distribución de medias muestrales	No hay diferencia en el método (pero se basa en la varianza poblacional estimada).
Forma de la distribución comparativa	Se utiliza la distribución *t* con $gl = N - 1$.
3. Determinar el punto de corte correspondiente al nivel de significación elegido.	Se utiliza la tabla *t* .
4. Determinar el valor muestral en la distribución comparativa.	No hay diferencia en el método (pero se denomina punto *t*).
5. Comparar los valores de los pasos 3 y 4 para decidir si se rechaza la hipótesis nula.	No hay diferencia en el método.

Otro ejemplo de prueba t con una sola muestra

Analicemos otro ejemplo ficticio. Supongamos que un investigador estaba estudiando los efectos psicológicos de una inundación devastadora en una pequeña comunidad rural. Específicamente, el investigador estaba interesado en saber si las personas se sentían más o menos esperanzadas después de la inundación. El investigador selecciona 10 personas al azar para que completen un pequeño cuestionario. El punto principal del cuestionario solicita a los individuos que clasifiquen en qué medida se sienten esperanzados, utilizando una escala de 7 puntos que va desde **extremadamente desesperanzado** (1), **pasando por neutro** (4), **hasta extremadamente esperanzado** (7). La tabla 9-3 muestra los resultados y cálculos de la prueba t para una sola muestra; la figura 9-4 representa gráficamente las distribuciones involucradas.

El investigador estaba interesado en saber si las respuestas estarían ubicadas consistentemente por encima o por debajo del punto medio de la escala (4). Los pasos de la prueba de hipótesis son los siguientes:

1. **Replantear el problema en función de hipótesis de investigación e hipótesis nula sobre poblaciones**. Existen dos poblaciones:

Población 1: personas que sufrieron la inundación.
Población 2: personas que no están ni esperanzadas ni desesperanzadas.

La hipótesis de investigación establece que las dos poblaciones producirán valores diferentes. La *hipótesis nula establece que producirán los mismos valores.*

2. **Determinar las características de la distribución comparativa**. Si la hipótesis nula es verdadera, la media de las dos distribuciones poblacionales será 4. Por otro lado, la varianza de estas distribuciones poblacionales se desconoce; debe ser estimada a partir de la muestra. Tal como lo indica la tabla 9-3, la suma de los desvíos cuadráticos con respecto a la media muestral es 32,10. Por lo tanto, la varianza poblacional estimada es 3,57, es decir, 32,10 dividido por 9 grados de libertad (10 – 1) es igual a 3,57.

La distribución de medias tendrá una media de 4 (igual a la media poblacional), su varianza es la varianza poblacional estimada dividida por el tamaño de la muestra, 3,57 dividido 10 es igual a 0,36. La raíz cuadrada de este resultado, es decir, el desvío estándar de la distribución de medias, es 0,60.

3. **Determinar el valor muestral de corte en la distribución comparativa, a partir del cual debería rechazarse la hipótesis nula**. El investigador desea ser muy cauteloso en cuanto a concluir equivocadamente que la inundación produjo una diferencia. Por lo tanto, decide probar la hipótesis al nivel 0,01. La hipótesis no era direccional (es decir, no se especificó una dirección determinada de la diferencia con respecto a la media de 4; cualquiera de los dos resultados habría sido de interés); por lo tanto, el investigador utiliza una prueba de dos colas y busca el punto de corte en la tabla 9-1 (o en la tabla B-2 del apéndice B), correspondiente a una prueba de dos colas y a 9 grados de libertad. El número que indica la tabla es 3,250. Por lo tanto, para rechazar la hipótesis nula el investigador necesita un t de 3,250 ó mayor, o bien un t de –3,250 ó menor.

4. **Determinar el valor muestral en la distribución comparativa**. La media muestral, que es de 4,7, se encuentra a 0,7 puntos de escala de la media de la hipótesis nula, que es de 4,0. La diferencia implica 1,17 desvíos estándar en la distribución comparativa con respecto a la media de esa distribución (0,7/0,6 = 1,17); $t = 1,17$.

5. **Comparar los valores de los pasos 3 y 4 para decidir si se rechaza o no la hipótesis nula.** El t de 1,17 no es tan extremo como el ±3,250 que se necesitaba para rechazar la hipótesis nula; por lo tanto, el investigador no puede rechazarla. El estudio no es concluyente. (Si el investigador hubiera utilizado una muestra más grande, que tuviera más potencia, el resultado podría haber sido bastante diferente).

Tabla 9-3.
Datos y análisis de una prueba *t* para una sola muestra referente a un estudio de la clasificación del nivel de esperanza de 10 individuos después de haber sufrido una inundación devastadora (datos ficticios).

	Clasificación	Diferencia con respecto a la media	Diferencia cuadrática con respecto a la media
	(X)	$(X-M)$	$(X-M)2$
	5	0,3	0,09
	3	-1,7	2,89
	6	1,3	1,69
	2	-2,7	7,29
	7	2,3	5,29
	6	1,3	1,69
	7	2,3	5,29
	4	-0,7	0,49
	2	-2,7	7,29
	5	0,3	0,09
Σ:	47	0	32,10

$M = \Sigma X/N = 47/10 = 4{,}7.$
$gl = N - 1 = 10 - 1 = 9.$
$\mu = 4{,}0.$
$S^2 = SC/gl = 32{,}10/(10-1) = 32{,}10/9 = 3{,}57.$
$S_M^2 = S2/N = 3{,}57/10 = 0{,}36.$
$S_M = \sqrt{S_M^2} = \sqrt{0{,}36} = 0{,}60.$
t necesaria para un nivel de significación del 1%, con $gl = 9$, en una prueba de dos colas $= \pm 3{,}250$.
t real de la muestra, $t = (M - \mu)/S_M = (4{,}7 - 4)/0{,}6 = 0{,}7/0{,}6 = 1{,}17.$

Decisión: no se rechaza la hipótesis nula.

Resumen de los pasos a seguir para realizar una prueba *t* para una sola muestra

La tabla 9-4 resume los pasos de la prueba de hipótesis cuando se trabaja con observaciones de una sola muestra y con una población de la cual se conoce la media pero no la varianza.

LA PRUEBA *t* PARA MEDIAS DEPENDIENTES

Hasta aquí hemos analizado ejemplos en los que conocemos la media de la población pero no la varianza. Este tipo de investigación es muy poco común. ¡Por lo general, uno no conoce siquiera la media poblacional! Ahora nos dedicaremos a una situación de investigación común en la que se desconocen la media y la varianza de la población. Este tipo de situación involucra estudios en los que se observan dos valores por cada una de las distintas personas. Por ejemplo, un psicólogo especializado en fisiología podría medir el patrón EEG (*Electroencephalogram*, Electroencefalograma) ("ondas cerebrales"), comparando el EEG de cada persona mientras realiza tareas abstractas en contraposición a cuando realiza tareas concretas. El tipo de investigación en el que cada persona es medida más de una vez se denomina **diseño de medidas repetidas**. (También se conoce como "diseño intra-sujeto". *Véase* en el apéndice A el resumen de las clases más importantes de diseños de investigación).

Figura 9-4. Distribuciones relacionadas con el ejemplo acerca del nivel de esperanza de individuos después de haber sufrido una inundación devastadora.

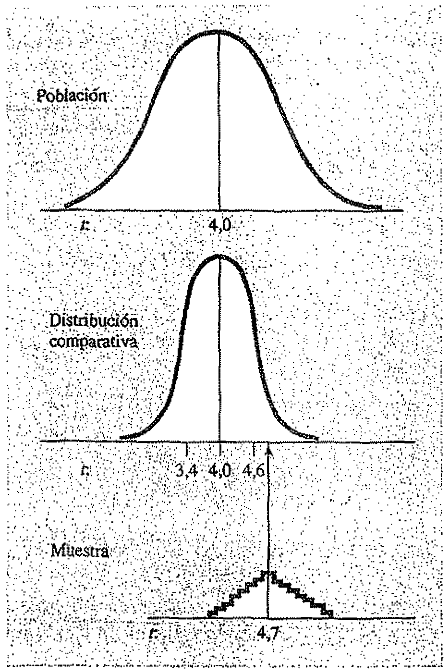


Tabla 9-4.
Pasos a seguir para realizar una prueba *t* para una sola muestra.

1. Replantear el problema en función de hipótesis de investigación e hipótesis nula sobre poblaciones.
2. Determinar las características de la distribución comparativa.
   a) La media es igual a la media poblacional conocida.
   b) El desvío estándar se calcula de la siguiente forma:
      i) Calcular la varianza poblacional estimada: $S^2 = SC/gl$.
      ii) Calcular la varianza de la distribución de medias: $S_M^2 = S^2/N$.
      iii) Calcular el desvío estándar: $S_M = \sqrt{S_M^2}$.
   c) La forma es la de una distribución t con $N - 1$ grados de libertad.
3. Determinar el valor muestral de corte en la distribución comparativa, a partir del cual se debería rechazar la hipótesis nula.
   a) Determinar los grados de libertad, el nivel de significación deseado, y la cantidad de colas de la prueba (una o dos).
   b) Buscar el punto de corte correspondiente en la tabla t.
4. Determinar el valor muestral en la distribución comparativa $t = (M - \mu)/S_M$.
5. Comparar los valores de los pasos 3 y 4 para decidir si se rechaza o no la hipótesis nula.

En uno de los diseños de medidas repetidas ampliamente utilizado se mide al mismo individuo antes y después de alguna intervención psicológica o social. Por ejemplo, un psicólogo empresarial podría medir los días de ausencia laboral de 80 trabajadores antes y después de la presentación de un nuevo programa de promoción sanitaria.

En esta situación común de diseño de medidas repetidas, en la que cada persona es medida dos veces, el procedimiento de prueba de hipótesis utilizado se denomina **prueba *t* para medias dependientes**. Se denominan "medias dependientes" porque las medias de cada grupo de valores (p. ej. valores anteriores y valores posteriores) dependen entre sí en cuanto pertenecen a la misma persona. (En el capítulo 10 analizamos el caso en el que un investigador compara valores de dos grupos diferentes de personas, es decir, un diseño de investigación analizado a través de una "prueba *t* para medias independientes").

La prueba *t* para medias dependientes es exactamente igual a la prueba *t* para una sola muestra, excepto que a) utilizamos algo llamado diferencias y b) suponemos que la media poblacional es 0. Veamos ahora cada uno de estos nuevos aspectos.

Diferencias

En un diseño de medidas repetidas, la muestra incluye dos valores por cada persona en lugar de uno sólo. Esto se maneja convirtiendo los dos valores por persona en uno sólo. El truco es crear **diferencias**: tomamos los valores de cada persona y restamos uno al otro.

Analicemos el ejemplo acerca del EEG. El psicólogo especializado en fisiología realizará una resta por cada persona: la medida del EEG de la persona durante la tarea abstracta menos la medida del EEG de la misma persona durante la tarea concreta. Así, se obtiene una sola diferencia abstracto-menos-concreto para cada persona. Similarmente, si tomamos el ejemplo de las ausencias laborales, el psicólogo empresarial realizará la siguiente resta por cada persona: la cantidad de días perdidos después del programa menos la cantidad de días perdidos antes del programa. El resultado sería una diferencia posterior-menos-anterior para cada empleado.

Cuando se trata de un valor anterior y de un valor posterior, generalmente tomamos el valor posterior y le restamos el anterior, para obtener una medida del cambio. En otros casos, tal como el ejemplo del EEG, realmente no importa cuál se resta a cuál, siempre que lo hagamos de la misma manera con todas las personas de la muestra.

Una vez que tenemos la diferencia de cada persona del estudio, realizamos el resto del procedimiento de prueba de hipótesis utilizando las diferencias. Es decir, procedemos como si se tratara de un estudio de una sola muestra de valores, los cuales, en este caso, resultan ser las diferencias.[2]

Población de diferencias con media 0

Hasta esta parte del libro, siempre hemos sabido cuál era la media de la población 2 (población con la que comparábamos la muestra). Por ejemplo, en la encuesta sobre las horas de estudio en el edificio de dormitorios de la facultad, sabíamos que la media poblacional de alumnos de la facul-

[2] También podemos utilizar una prueba *t* para medias dependientes en una situación en la que tenemos valores de pares de participantes en la investigación. Analizamos cada par como si fuera una persona y calculamos una diferencia por cada par. Por ejemplo, supongamos que tenemos 30 parejas de matrimonios y estamos comparando edades de esposos y esposas para ver si los esposos son sostenidamente mayores que las esposas. Podríamos calcular para cada pareja una diferencia de la edad del esposo menos la de la esposa. Luego realizaríamos el resto de la prueba de hipótesis del mismo modo que cualquier otra prueba *t* para medias dependientes. Cuando la prueba *t* para medias dependientes se utiliza de este modo, a veces se la llama **prueba *t* para diseños apareados** o **prueba *t* de comparaciones pareadas**.

tad era, en general, 2,5 horas. Sin embargo, ahora estamos utilizando diferencias, y por lo general no conocemos la media poblacional de las mismas.

La solución es la siguiente: comúnmente, la hipótesis nula en un diseño de medidas repetidas establece que no hay diferencia entre los dos grupos de valores. Por ejemplo, la hipótesis nula del estudio realizado por el psicólogo especializado en fisiología es que la actividad EEG será la misma al hacer tareas abstractas o concretas. Similarmente, la hipótesis nula del estudio acerca de la promoción sanitaria establece que las inasistencias laborales serán iguales antes y después de presentar el programa de promoción sanitaria. Por lo tanto, al utilizar diferencias usualmente comparamos una hipótesis de investigación que establece una diferencia predicha, con una hipótesis nula que establece una diferencia nula.

El punto clave es el siguiente: ¿Qué significa "diferencia nula"? Es decir, ¿qué significa decir que en la población, en líneas generales, la diferencia entre los dos valores de una persona es nula? Es lo mismo que decir que la media de la población de diferencias es 0. En otras palabras, decir que la diferencia entre los dos valores es nula es equivalente a decir que el promedio de las diferencias es cero.

Por lo tanto, al trabajar con diferencias suponemos una población comparativa artificial de diferencias que tiene una media poblacional igual a 0.

Ejemplo de prueba *t* para medias dependientes

Olthoff (1989) analizó la calidad de la comunicación entre parejas comprometidas tres meses antes y tres meses después del matrimonio. Uno de los grupos estudiados estaba formado por 19 parejas que habían recibido el acostumbrado curso prematrimonial por parte de los ministros que iban a celebrar su matrimonio. (Para que el ejemplo no se complique, nos concentraremos sólo en este grupo, y únicamente en los esposos que forman el grupo. Los valores de las esposas eran similares, aunque un poco más variados, haciéndolos algo más complicados como ejemplo para aprender el procedimiento de la prueba *t*).

Los valores de los 19 esposos están enumerados en las columnas "Antes" y "Después" de la tabla 9-5, seguidas del análisis completo de la prueba *t*. (Las distribuciones involucradas aparecen en la figura 9-5). La media de los valores anteriores fue 116,316 y la media de los valores posteriores fue 104,263. Lo más importante, sin embargo, son las diferencias, que también hemos calculado. La media de los registros diferenciales es –12,05. En promedio, la calidad de comunicación de estos maridos disminuyó aproximadamente 12 puntos.

¿Es significativa esta disminución? En otras palabras, ¿cuán probable es que esta muestra de valores de cambio sea una muestra aleatoria de una población de valores de cambio cuya media es 0? Realicemos el procedimiento de la prueba de hipótesis.

1. **Replantear el problema en función de hipótesis de investigación e hipótesis nula de las poblaciones.** Las dos poblaciones son:

Población 1: maridos que asisten al curso prematrimonial acostumbrado.

Población 2: maridos cuya calidad de comunicación anterior al matrimonio no cambia después de casados.

La hipótesis de investigación establece que la población 1 es diferente de la población 2, es decir, los maridos que asisten al acostumbrado curso prematrimonial (tal como los maridos que analizó Olthoff) sí cambian en cuanto a calidad de comunicación antes y después del matrimonio. La hipótesis nula establece que las poblaciones son iguales, que los maridos que asisten al acostumbrado curso prematrimonial **no** cambian en cuanto a la calidad de su comunicación antes y después del matrimonio.

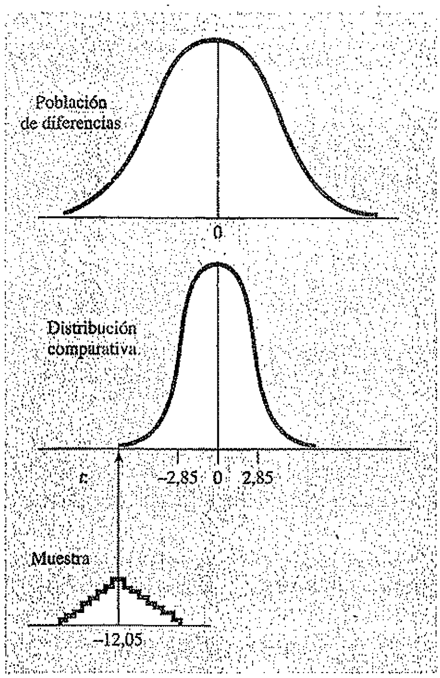


Figura 9-5. Distribuciones relacionadas con el ejemplo de Olthoff (1993) de una prueba t para medias dependientes.

Es importante destacar que no tenemos información real acerca de los maridos de la población 2. Los maridos del estudio son una muestra de la población 1 de maridos. Si la hipótesis de investigación es correcta, es probable que los maridos de la población 2 ni siquiera existan. Sólo con el propósito de realizar la prueba de hipótesis, establecimos la población 2 como una especie de grupo comparativo de hombres en pareja. Es decir, establecimos un grupo comparativo con el propósito de analizar maridos que, si se miden antes y después del matrimonio, no mostrarían ningún cambio.

2. **Determinar las características de la distribución comparativa.** Si la hipótesis nula es verdadera, la media poblacional de las diferencias es 0. La varianza poblacional de las diferencias puede estimarse a partir de la muestra de las diferencias. Tal como lo indica la tabla 9-5, la suma de los desvíos cuadráticos de las diferencias con respecto a la media de diferencias es 2.772,9. Al haber 19 maridos en el estudio, existen 18 grados de libertad. Dividiendo la suma de los desvíos cuadráticos por los grados de libertad, obtenemos una varianza poblacional estimada de 154,05.

La distribución de medias (de esta población de diferencias) tendrá una media de 0, al igual que la media poblacional; su varianza será la varianza poblacional estimada (154,05) dividida por el tamaño de la muestra (19), lo que da 8,11. El desvío estándar es la raíz cuadrada de 8,11, que es 2,85. Dado que Olthoff estaba utilizando una varianza poblacional estimada, la distribución comparativa es una distribución t. La estimación de la varianza poblacional se realizó sobre la base de 18 grados de libertad, por lo tanto, esta distribución comparativa es una distribución t para 18 grados de libertad.

3. **Determinar el punto de corte en la distribución comparativa, apartir del cual debería rechazarse la hipótesis nula.** Olthoff utilizó una prueba de dos colas porque no existía razón evi-

Tabla 9-5.
Análisis de la prueba t referida a los registros de calidad de comunicación antes y después del matrimonio, realizado a 19 esposos que no recibieron ninguna capacitación especial en cuanto a comunicación.

Esposo	**Calidad de comunicación** *Antes*	*Después*	**Diferencia (después – antes)**	**Desvío de las diferencias con respecto a la media diferencial**	**Desvío cuadrático**
A	126	115	–11	1,05	1,1
B	133	125	–8	4,05	16,4
C	126	96	–30	–17,95	322,2
D	115	115	0	12,05	145,2
E	108	119	11	23,05	531,3
F	109	82	–27	–14,95	233,5
G	124	93	–31	–18,95	359,1
H	98	109	11	23,05	531,3
I	95	72	–23	–10,95	119,9
J	120	104	–16	–3,95	15,6
K	118	107	–11	1,05	1,1
L	126	118	–8	4,05	16,4
M	121	102	–19	–6,95	48,3
N	116	115	–1	11,05	122,1
O	94	83	–11	1,05	1,1
P	105	87	–18	–5,95	35,4
Q	123	121	–2	10,05	101,0
R	125	100	–25	–12,95	167,7
S	128	118	–10	2,05	4,2
Σ:	2.210	1.981	–229		2.772,9

Para las diferencias:

$M = -229/19 = -12{,}05$.

$\mu = 0$ (tomado como base comparativa de ausencia de cambio).

$S^2 = SC/gl = 2.772{,}9/(19 - 1) = 154{,}05$.

$S_M^2 = S^2/N = 154{,}05/19 = 8{,}11$.

$S_M = \sqrt{S_M^2} = \sqrt{8{,}11} = 2{,}85$.

t necesario para el nivel 5%, con $gl = 18$ y prueba de dos colas $= \pm 2{,}101$.

$t = (M - \mu)/S_M = (-12{,}05 - 0)/2{,}85 = -4{,}23$.

Decisión: se rechaza la hipótesis nula

Fuente: Olthoff (1989).

dente para predecir un aumento o una disminución en la calidad de la comunicación. La tabla B-2 indica que utilizando un nivel de significación de 0,05 y 18 grados de libertad para rechazar la hipótesis nula, se necesita un punto t de +2,101 ó mayor, o bien de –2,101 ó menor.

4. **Determinar el valor muestral en la distribución comparativa.** La muestra de Olthoff tenía una media de diferencias de –12,05. Es decir, la media estaba 12,05 puntos por debajo de la media de distribución de medias, que es igual a 0. El desvío estándar de la distribución de medias que calculamos era de 2,85. Por lo tanto, la media de las diferencias –12,05 se encuentra 4,23 desvíos estándar por debajo de la media de la distribución de medias, es decir, la muestra de diferencias de Olthoff corresponde a un punto t de – 4,23

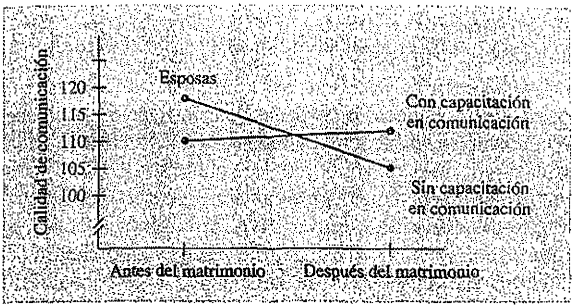


Figura 9-6. Capacidad de comunicación de esposas que reciben capacitación prematrimonial sobre comunicación y de esposas que no reciben dicha capacitación (sobre la base de Olthoff, 1989).

5. **Comparar los valores de los pasos 3 y 4 para decidir si se rechaza o no la hipótesis nula.** El t de $-4{,}23$ de la muestra de diferencias es más extremo que el t $\pm 2{,}101$ necesario. Por lo tanto, podemos rechazar la hipótesis nula. El resultado sugiere que los maridos analizados por Olthoff pertenecen a una población en la que la calidad de comunicación de los maridos después del matrimonio es diferente de lo que era antes (es menor).

El estudio real de Olthoff era más complejo. Tal vez resulte interesante saber que se descubrió que las esposas también mostraban esta disminución en cuanto a calidad de comunicación después de casadas. Sin embargo, un grupo similar de parejas comprometidas, a quienes sus ministros dieron capacitación especial sobre capacidad de comunicación (mucho mayor que la acostumbrada sesión breve) no mostraron una disminución significativa en la calidad de comunicación marital después del matrimonio (*véase* la figura 9-6). De hecho, actualmente existe gran cantidad de investigación que indica que la calidad marital de todo tipo disminuye en líneas generales (p. ej., Karney & Bradbury), y los cursos intensivos sobre capacidad de comunicación pueden ser muy útiles para reducir o eliminar esta disminución (Markman et al., 1993).

Otro ejemplo de prueba t para medias dependientes

Aquí tenemos otro ejemplo. Un investigador está interesado en el efecto producido por el ruido en la coordinación entre el pulso y la vista de los cirujanos. El investigador toma una prueba estándar de coordinación entre el pulso y la vista a nueve cirujanos en ambas condiciones, silenciosa y ruidosa (no mientras operaban, por supuesto). La predicción era que la coordinación de los cirujanos es mayor en condiciones de silencio. (Lo ideal sería que cualquier efecto que pudiera producir práctica o fatiga, por realizar dos veces la prueba de coordinación entre el pulso y la vista, sea equiparado poniendo a prueba primero una mitad de los cirujanos en condiciones de ruido, y la otra mitad, también primero, en condiciones de silencio. El apéndice A describe este diseño contrabalanceado).

La tabla 9-6 indica los resultados de este estudio ficticio. También muestra el cálculo de las diferencias y todos los otros cálculos de la prueba t para medias dependientes. La figura 9-7 representa gráficamente las distribuciones involucradas. Los siguientes son los pasos de la prueba de hipótesis:

1. **Replantear el problema en función de hipótesis de investigación e hipótesis nula de las poblaciones.** Las dos poblaciones son:

Población 1: cirujanos como *los analizados en el estudio.*
Población 2: cirujanos cuya coordinación es la misma en condiciones de silencio o ruido.

La hipótesis de investigación establece que la media de las diferencias de la población 1 (silencioso menos ruidoso) es mayor que la de la población 2. Es decir, la hipótesis de investigación establece que los cirujanos se desempeñan mejor en condiciones de silencio. La hipótesis nula establece que la diferencia, en cuanto a desempeño de la población 1, no es mayor que la de la población 2. Es decir, la hipótesis nula establece que los cirujanos no se desempeñan mejor en condiciones de silencio.

2. **Determinar las características de la distribución comparativa.** Si la hipótesis nula es verdadera, la media poblacional de las diferencias es 0. ¿Cuál es la varianza de esta población de diferencias? Estimándola a partir de la muestra de diferencias, es la suma de los desvíos cuadráticos de las diferencias con respecto a su media, dividida por los grados de libertad. El resultado aparece en la tabla 9-6 y es igual a 7,5. La distribución comparativa es una distribución de medias; la varianza es la varianza de la distribución de observaciones individuales (en este caso una varianza estimada) dividida por el tamaño de la muestra: 7,5/9 = 0,83. El desvío estándar de la distribución de medias es 0,91 (la raíz cuadrada de 0,83). La forma de la distribución comparativa será una distribución t con 8 grados de libertad.

3. **Determinar el punto de corte en la distribución comparativa, a partir del cual debería rechazarse la hipótesis nula.** Estamos trabajando con una prueba de una cola porque había una base razonable para predecir la dirección de la diferencia. Supondremos que el investigador quería ser conservador y utilizó un nivel de significación del 1%. Con 8 grados de libertad, la tabla B-2 indica que es necesario un punto t de al menos 2,897 para rechazar la hipótesis nula.

4. **Determinar el valor muestral en la distribución comparativa.** La diferencia media de la muestra, que es igual a 2, se ubica 2,20 desvíos estándar (de 0,91) por encima de la media de la distribución de medias, que es igual a 0.

5. **Comparar los valores de los pasos 3 y 4 para decidir si se rechaza o no la hipótesis nula.** El valor t muestral de 2,20 es menos extremo que el punto de corte t de 2,897. Por lo tanto, no se puede rechazar la hipótesis nula. El experimento no es concluyente. (A propósito, si el investigador hubiera establecido el nivel de significación en 0,05, el resultado hubiera sido significativo).

Tercer ejemplo de prueba t para medias dependientes

Un psicólogo especializado en desarrollo está estudiando la sensibilidad de los niños frente a extraños, utilizando un nuevo tipo de medida. Tiene la posibilidad de medir a 10 niños a los 3 meses de edad y hacerlo nuevamente a los 4 meses. Su predicción es que habrá un aumento de sensibilidad.

La tabla 9-7 indica los resultados de este estudio ficticio, junto con el cálculo de las diferencias y todos los otros cálculos de la prueba t para medias dependientes. La figura 9-8 representa gráficamente las distribuciones involucradas. Los pasos de la prueba de hipótesis son los siguientes.

1. **Replantear el problema en función de hipótesis de investigación e hipótesis nula de las poblaciones.** Las dos poblaciones son:

Población 1: niños como los analizados en este estudio.
Población 2: niños cuya sensibilidad hacia extraños es la misma a los 3 meses que a los 4 meses de edad.

La hipótesis de investigación establece que la media de las diferencias de la población 1 (sensibilidad hacia extraños a los 4 meses menos sensibilidad a los 3 meses) es mayor que la de la población 2. La hipótesis nula establece que la media de las diferencias de la población 1 no es mayor que la de la población 2.

Tabla 9-6.
Prueba *t* de un estudio acerca de la coordinación entre el pulso y la vista, en el que se mide a nueve cirujanos en condiciones de silencio y ruido (datos ficticios).

Cirujano	Condiciones		Diferencia	Desvío	Desvío Cuadrático
	Silencio	Ruido			
1	18	12	6	6 – 2 = 4	16
2	21	21	0	–2	4
3	19	16	3	1	1
4	21	16	5	3	9
5	17	19	–2	–4	16
6	20	19	1	–1	1
7	18	16	2	0	0
8	16	17	–1	–3	9
9	20	16	4	2	4
S:	170	152	18	0	60

Para las diferencias:

$M = 18/9 = 2{,}0.$

$\mu = 0$ (tomado como base comparativa de ausencia de cambio).

$S^2 = SC/gl = 60/(9-1) = 60/8 = 7{,}5.$

$S_M^2 = S^2/N = 7{,}50/9 = 0{,}83.$

$S_M = \sqrt{S_M^2} = \sqrt{0{,}83} = 0{,}91.$

t necesario para un nivel de significación del 1%, $gl = 8$ y prueba de una cola = 2,897.

$t = (M - \mu)/S_M = (2{,}00 - 0)/0{,}91 = 2{,}20.$

Decisión: no se rechaza la hipótesis nula..

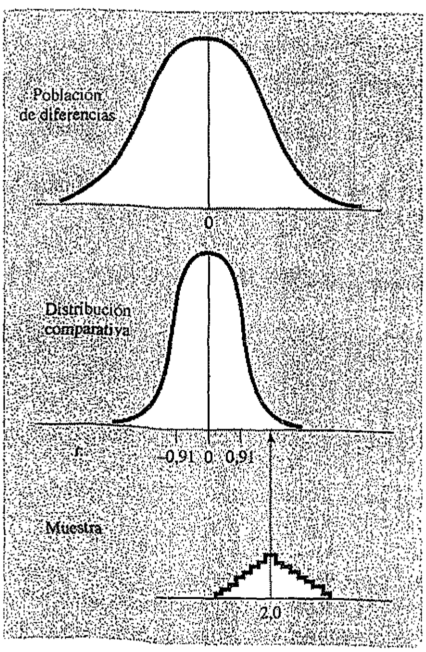


Figura 9-7. Distribuciones relacionadas con el estudio ficticio acerca de la coordinación entre el pulso y la vista en condiciones de ruido y silencio.

2. **Determinar las características de la distribución comparativa.** La media poblacional es una diferencia igual a 0. La varianza poblacional estimada, según lo indica la tabla 9-7, es igual a 0,39. La distribución comparativa será una distribución *t* con 9 grados de libertad, una media de 0 y un desvío estándar de 0,20.

3. **Determinar el punto muestral de corte en la distribución comparativa, a partir del cual debería rechazarse la hipótesis nula.** Estamos trabajando con una prueba de una cola (porque existía base razonable para predecir la dirección de la diferencia). La tabla B-2 indica que utilizando un nivel de significación del 5% y 9 grados de libertad, es necesario un punto *t* de al menos 1,833 para rechazar la hipótesis nula.

4. **Determinar el valor muestral en la distribución comparativa.** El cambio de la media muestral, que es igual a 0,14, se ubica 0,70 desvíos estándar (de 0,20 cada uno) por encima de la media de la distribución de medias, que es igual a 0.

5. **Comparar los valores de los pasos 3 y 4 para decidir si se rechaza o no la hipótesis nula.** El *t* muestral, que es igual a 0,70, es menos extremo que el *t* necesario de 1,833. Por lo tanto, no se puede rechazar la hipótesis nula. El estudio no es concluyente.

Resumen de los pasos a seguir para la realización de una prueba *t* para muestras dependientes

La tabla 9-8 resume los pasos para la realización de la prueba *t* para medias dependientes. El apéndice del capítulo incluye las fórmulas de cálculo optativas para facilitar la realización de la prueba *t* para medias dependientes en forma manual, cuando se trabaja con una gran cantidad de diferencias.

SUPUESTOS DE LA PRUEBA *T*

Como ya hemos visto, al utilizar una varianza poblacional estimada, la distribución comparativa es una distribución *t*. Sin embargo, la distribución comparativa será exactamente una distribución *t* sólo si la distribución de observaciones individuales sigue una distribución normal. De lo contrario, la distribución comparativa tendrá alguna otra forma (generalmente desconocida).

Por lo tanto, hablando estrictamente, una población normal es condición necesaria dentro de la lógica y de la matemática de una prueba *t*. Una condición de este tipo en un procedimiento de prueba de hipótesis se denomina **supuesto**. Se dice que una distribución poblacional normal es un supuesto de la prueba *t*. El efecto de este supuesto es que si la distribución poblacional no es normal, es técnicamente incorrecto utilizar la prueba *t*.

Lamentablemente, por lo general no sabemos si la población es normal, ya que cuando realizamos una prueba *t*, usualmente todo lo que tenemos para trabajar son los valores muestrales. Afortunadamente, como vimos en el capítulo 5, las distribuciones en las investigaciones psicológicas con mucha frecuencia se aproximan a la curva normal. (Esto también se aplica a las distribuciones de diferencias). Además, los estadísticos han descubierto que, en la práctica, aun cuando la población se encuentre bastante lejos de lo normal, con la prueba *t* se obtienen resultados razonablemente precisos. En otras palabras, se dice que la prueba *t* es **robusta** más allá de incumplimientos moderados del supuesto de una distribución poblacional normal. Es interesante la forma en que los estadísticos calculan la **robustez** de una prueba, tema que describiremos en el cuadro 10-1 del capítulo 10.

Existe una situación razonablemente común en la que utilizar una prueba *t* para medias dependientes puede dar resultados seriamente distorsionados. Es el caso en el que realizamos una prueba de una cola y la población es muy asimétrica (con una de las colas mucho más larga que la otra).

Tabla 9-7.
Prueba t de un estudio acerca de la sensibilidad de 10 niños hacia los extraños, medida a los 3 y a los 4 meses de edad (datos ficticios).

Niño	Edad		Diferencia	Desvío	Desvío Cuadrático
	3 meses	4 meses			
1	10,4	10,8	0,4	0,26	0,07
2	12,6	12,1	–0,5	–0,64	0,41
3	11,2	12,1	0,9	0,76	0,58
4	10,9	11,4	0,5	0,36	0,13
5	14,3	13,9	–0,4	–0,54	0,29
6	13,2	13,5	0,3	0,16	0,03
7	9,7	10,9	1,2	1,06	1,12
8	11,5	11,5	0,0	–0,14	0,02
9	10,8	10,4	–0,4	–0,54	0,29
10	13,1	12,5	–0,6	–0,74	0,55
S:	117,70	119,10	1,4	0	3,49

Para las diferencias:

$M = 1{,}4/10 = 0{,}14.$

$\mu = 0.$

$S^2 = SC/gl = 3{,}49/(10 - 1) = 3{,}49/9 = 0{,}39.$

$S_M^2 = S^2/N = 0{,}39/10 = 0{,}039.$

$S_M = \sqrt{S_M^2} = \sqrt{0{,}039} = 0{,}20.$

t necesario para el nivel de significación 5%, $gl = 9$ y prueba de una cola = 1,833.

$t = (M - \mu)/S_M = (0{,}14 - 0)/0{,}20 = 0{,}70.$

Decisión: no se rechaza la hipótesis nula..

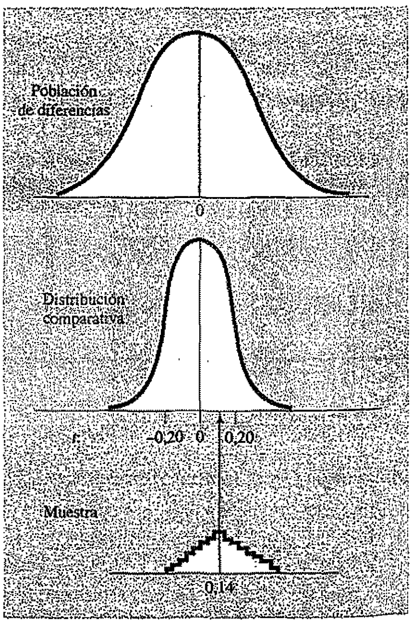


Figura 9-8. Distribuciones relacionadas con un estudio ficticio acerca de la sensibilidad de niños hacia extraños, a los 3 y a los 4 meses de edad.

Tabla 9-8.
Pasos para la realización de una prueba t para medias dependientes.

1. Replantear el problema en función de hipótesis de investigación e hipótesis nula de las poblaciones.
2. Determinar las características de la distribución comparativa.
   a) Convertir los dos valores de cada persona en una diferencia. Realizar todos los pasos restantes utilizando las diferencias.
   b) Calcular la media de las diferencias.
   c) Presumir una media poblacional igual a 0: $\mu = 0$.
   d) Calcular la varianza poblacional estimada de diferencias $S^2 = SC/gl$.
   e) Calcular la varianza de la distribución de medias de diferencias: $S_M^2 = S^2/N$.
   f) Calcular el desvío estándar de la distribución de medias de las diferencias: $S_M = \sqrt{S_M^2}$
   g) La forma es la de una distribución t con $gl = N - 1$.
3. Determinar el punto de corte en la distribución comparativa, a partir del cual debería rechazarse la hipótesis nula.
   a) Determinar el nivel de significación deseado y si se utilizará una prueba de una o dos colas.
   b) Buscar el punto de corte indicado en una tabla t.
4. Determinar el valor muestral en la distribución comparativa: $t = (M - \mu)/S_M$.
5. Comparar los valores de los pasos 3 y 4 para decidir si se rechaza o no la hipótesis nula.

¿Cómo sabemos que la población es muy asimétrica? Un caso puede ser aquel en el que la muestra de diferencias es muy asimétrica. Si la muestra es muy asimétrica, es probable que la población de donde proviene la muestra sea muy asimétrica también. Otro caso es aquel en el que existen razones para pensar que se produce un efecto techo o piso que hace que la distribución sea asimétrica porque los valores de un lado no pueden ser mayores o menores a determinado punto. Existen varias alternativas para reemplazar la prueba t, cuando hay razones para creer que realizarla violaría seriamente el supuesto de normalidad y daría resultados distorsionados. En el capítulo 15 veremos esas alternativas.

TAMAÑO DEL EFECTO Y POTENCIA DE LA PRUEBA t PARA MEDIAS DEPENDIENTES

Tamaño del efecto

El tamaño del efecto, en un estudio en el que se utiliza una prueba t para medias dependientes, se calcula del mismo modo que en el capítulo 8. Es la diferencia entre las medias poblacionales dividida por el desvío estándar de la población: $(\mu_1 - \mu_2)/\sigma$. Sin embargo, al utilizar diferencias, la media de la población 2 usualmente es 0 (es decir, cuando se trabaja con diferencias, $\mu_{2s} = 0$). Esto simplifica la situación:

$$d = \frac{(\mu_1 - 0)}{\sigma} = \frac{\mu_1}{\sigma} \tag{9-8}$$

Es importante recordar que cuando se utiliza esta fórmula, μ_1 es la media predicha de la población de diferencias y σ es el desvío estándar de las poblaciones de diferencias.

Las reglas del tamaño del efecto de una prueba *t* para medias dependientes son las mismas que aprendimos para el caso analizado en el capítulo 8: un tamaño del efecto pequeño es igual a 0,20, uno mediano es igual a 0,50, y uno grande es igual a 0,80.

Analicemos un ejemplo. Un psicólogo especializado en deportes planifica un estudio acerca de las actitudes hacia compañeros de equipo antes y después del juego. Realizará un cuestionario sobre actitudes dos veces, una antes y otra después del juego. Supongamos que la diferencia mínima entre antes y después, que puede tener cierta importancia, es de 4 puntos del cuestionario. Supongamos además que sobre la base de investigaciones relacionadas con el tema, el investigador calcula que el desvío estándar de las diferencias del cuestionario de actitud es aproximadamente de 8 puntos. Así, $\mu_1 = 4$ y $\sigma = 8$. Aplicando la fórmula para calcular el tamaño del efecto, $d = \mu_1/\sigma = 4/8 = 0{,}50$. Conforme a las reglas del tamaño del efecto, el estudio planificado tiene un tamaño del efecto mediano.

Si deseáramos estimar el tamaño del efecto después de haber realizado el estudio, dividiríamos la media real de las diferencias de la muestra por el desvío estándar estimado de la población de diferencias.

$$d = \frac{M}{S} \tag{9-9}$$

Es importante recordar que, en ésta fórmula, tanto *M* como *S* se refieren a diferencias. Además, *S* es el desvío estándar de la población de observaciones individuales (es decir, en este caso, de las diferencias de los individuos). No es lo mismo que S_M, el desvío estándar de la distribución de medias (de diferencias).

Analicemos nuestro primer ejemplo de prueba *t* para medias dependientes, el estudio acerca del cambio de los maridos en cuanto a la calidad de la comunicación. En ese estudio, la media de las diferencias era –12,05, y el desvío estándar poblacional estimado de diferencias sería 12,41. Es decir, calculamos la varianza estimada de registros diferenciales (S^2) y nos da 154,05; $\sqrt{S^2} = 12{,}41$. Por lo tanto, el tamaño de efecto se calcula como $d = M/S = -12{,}05/12{,}41 = -0{,}97$. Se trata de un tamaño del efecto muy grande. (El signo negativo del tamaño del efecto significa que el gran efecto era una disminución).

Potencia

La tabla 9-9 indica la potencia aproximada a un nivel de significación de 0,05 para los tamaños del efecto pequeños, medianos y grandes, correspondientes a pruebas de una o dos colas. En el ejemplo del psicólogo especializado en deportes, el investigador esperaba un tamaño del efecto mediano ($d = 0{,}50$). Si planificara realizar un estudio utilizando el nivel 0,05, con una prueba de dos colas y con 20 participantes, el estudio tendría una potencia de 0,59. Lo cual significa que si la hipótesis de investigación es realmente verdadera y tiene un tamaño del efecto mediano, existe un 59% de chances de que el estudio resulte significativo.

La tabla de potencia (tabla 9-9) también es útil cuando leemos el resultado no significativo de algún estudio publicado. Supongamos que un estudio que utiliza una prueba *t* para medias dependientes tuviera un resultado no significativo. El estudio probó la significación al nivel 0,05, con

Tabla 9-9.
Potencia aproximada de estudios en los que se utiliza la prueba t para medias dependientes en pruebas de hipótesis con nivel de significación de 0,05.

Registros de diferencias de la muestra (N)	**Tamaño de efecto**		
	Pequeño *(d = 0,20)*	*Mediano* *(d = 0,50)*	*Grande* *(d = 0,80)*
Prueba de dos colas			
10	0,09	0,32	0,66
20	0,14	0,59	0,93
30	0,19	0,77	0,99
40	0,24	0,88	*
50	0,29	0,94	*
100	0,25	*	*
Prueba de una cola			
10	0,15	0,46	0,78
20	0,22	0,71	0,96
30	0,29	0,86	*
40	0,35	0,93	*
50	0,40	0,97	*
100	0,63	*	*

*La potencia es casi 1.

una prueba de dos colas, y contaba con 10 participantes. ¿Deberíamos concluir que, en efecto, no existe ninguna diferencia entre las poblaciones? Probablemente no. Aun suponiendo un tamaño de efecto mediano, la tabla 9-9 indica que existe sólo un 32% de chances de obtener un resultado significativo en este estudio. Analicemos ahora otro estudio que resultó no significativo, en el que también se utilizó el nivel de significación 0,05 y una prueba de dos colas, pero que contaba con 100 participantes. La tabla 9-9 indica que existiría un 63% de chances de que el estudio resultara significativo si existiera incluso un tamaño del efecto real pequeño en la población. Si en la población hubiera un tamaño del efecto mediano, la tabla indica que existiría casi un 100% de chances de que el estudio resultase significativo. Por lo tanto, en este estudio con 100 participantes podríamos concluir, a partir de los resultados, que en la población probablemente no existe ninguna diferencia o que, en el mejor de los casos, existe una muy pequeña.

Para que la tabla 9-9 resultara simple, hemos incluido sólo la potencia correspondiente a unas pocas cantidades diferentes de participantes (10, 20, 30, 40, 50 y 100). Estos datos deberían ser suficientes para el tipo de evaluaciones aproximadas que se realizan al analizar resultados de publicaciones científicas.[3]

[3] Cohen (1988, pp. 28-39) proporciona tablas más detalladas en cuanto a cantidades de participantes, niveles de tamaño del efecto y niveles de significación. Si se utilizan sus tablas, debe tenerse en cuenta que la d a la que se hace referencia está basada en realidad en una prueba t para medias independientes (que es la situación que trataremos en el capítulo 10). Para utilizar esas tablas para una prueba t para medias dependientes, primero se debe multiplicar el tamaño del efecto deseado por 1,4. Por ejemplo, si el tamaño del efecto es 0,30, para utilizar las tablas de Cohen consideraríamos que es de 0,42 (es decir, $0{,}30 \times 1{,}4 = 0{,}42$). La única otra diferencia con respecto a nuestra tabla es que Cohen describe el nivel de significación con la letra a (por "nivel alfa"), con un subíndice de 1 ó 2, haciendo referencia a una prueba de una o dos colas. Por ejemplo, una tabla que en la parte superior indica $a1 = 0{,}05$ significa que es una tabla para $p < 0{,}05$, con una cola.

Planificación del tamaño de la muestra

La tabla 9-10 indica la cantidad aproximada de participantes necesarios para tener un 80% de potencia con tamaños de efecto pequeños, medianos o grandes, utilizando pruebas de una o dos colas con nivel de significación de 0,05. (Ochenta por ciento es un número comúnmente utilizado por los investigadores como potencia mínima para que tenga sentido realizar un estudio). Supongamos que planificamos un estudio en el que esperamos tener una gran tamaño del efecto y utilizáramos un nivel de significación de 0,05, con dos colas. La tabla indica que sólo necesitaríamos 14 participantes para tener una potencia del 80%. Por otro lado, un estudio en el que se utiliza el mismo nivel de significación y en el que se realiza una prueba de dos colas, pero en el que se espera sólo un pequeño tamaño del efecto, necesitaríamos 196 participantes para tener una potencia del 80%.[4]

La potencia de estudios en los que se emplea la prueba *t* para medias dependientes

Los estudios que utilizan diferencias (es decir, estudios que emplean el diseño de medidas repetidas) presentan con frecuencia tamaños del efecto considerablemente mayores a los de otros tipos de diseños de investigación con la misma cantidad de diferencia esperada entre las medias. Si los tamaños del efecto son mayores, entonces la potencia es mayor. Es decir, probar a cada uno de los participantes de un grupo dos veces (una vez en unas condiciones y otra vez en otras condiciones) usualmente da como resultado un estudio con alta potencia. En particular, este tipo de estudio da más potencia que dividir a los participantes en dos grupos y probar una vez a cada grupo (un grupo en unas condiciones y el otro grupo en las otras condiciones). De hecho, los estudios en los que se utilizan diferencias generalmente tienen más potencia que aquellos en los que se utiliza el doble de participantes probados sólo una vez cada uno.

¿Por qué los diseños de medidas repetidas tienen tanta potencia? La razón es que el desvío estándar de las diferencias generalmente es bastante bajo. (El desvío estándar de las diferencias es el valor por el cual realizamos la división para obtener el tamaño del efecto cuando utilizamos diferencias). En un diseño de medidas repetidas, la única variación es la de las diferencias. La variación entre participantes en cuanto a los valores en cada una de las prueba no forma parte de la variación involucrada en el análisis, ya que las diferencias comparan a los participantes consigo mismos. William S. Gosset fue, en esencia, el inventor de la prueba *t* (*véase* cuadro 9-1). Éste supo aprovechar el mayor nivel de potencia de los estudios con medidas repetidas en una controversia históricamente interesante acerca de un experimento relacionado con la leche, el cual se describe en el cuadro 9-2.

Tabla 9-9.
Cantidad aproximada de participantes necesarios para lograr un 80% de potencia en la prueba *t* para medias dependientes, en pruebas de hipótesis con un nivel de significación del 0,05.

	Tamaño del efecto		
	Pequeño ($d = 0{,}20$)	Mediano ($d = 0{,}50$)	Grande ($d = 0{,}80$)
Dos colas	196	33	14
Una cola	156	26	12

[4] Cohen (1988, pp. 54-55) proporciona tablas más detalladas que indican la cantidad necesaria de participantes para otros niveles de potencia además del de 80% (y también para otros tamaños del efecto además de las de 0,20, 0,50 y 0,80, así como también para otros niveles de significación). De todos modos, para la utilización de esas tablas se deben tener en cuenta las mismas indicaciones que en la nota al pie N° 3).

Cuadro 9-2. La potencia en estudios en los que se utilizan diferencias: cómo el experimento de Lanarkshire acerca del consumo de leche podría haber sido mejor aprovechado.

En el año 1930, se realizó en Escocia un importante experimento sanitario que involucraba a 20.000 alumnos. Su principal objetivo era comparar el crecimiento de un grupo de niños, a quienes se les hacía beber leche regularmente, con el de otros niños que formaban parte del grupo control. Los resultados obtenidos indicaron que aquellos que tomaban leche mostraban un crecimiento mayor.

Sin embargo, William S. Gosset, un estadístico de la época (*véase* cuadro 9-1), estaba asombrado por la manera en que se realizó el experimento. ¡Había costado £ 7.500, lo que en 1930 era una inmensa cantidad de dinero, y se había realizado erróneamente! Los grandes estudios del estilo del que tratamos eran muy populares entre los estadísticos de la época porque parecían imitar las grandes cantidades que se encuentran en la naturaleza. Gosset, por el contrario, siendo fabricante de cerveza, estaba obligado a utilizar en sus estudios cantidades muy reducidas, las tandas experimentales de cerveza eran muy costosas y, con frecuencia, era reprendido por los "verdaderos estadísticos" debido a los pequeños tamaños de muestra que utilizaba. No obstante, Gosset sostenía que ninguna cantidad de participantes era lo suficientemente grande cuando no se realizaba una asignación estrictamente aleatoria. ¡Y en el estudio mencionado, se permitió a los maestros intercambiar a los niños de un grupo a otro si sentían pena por alguno que creían que podía beneficiarse más al recibir la leche! (*Véase* en el apéndice A una exposición acerca de la asignación aleatoria de los participantes a cada grupo).

De todos modos, es aún más interesante, en vista de lo aprendido en este capítulo, que Gosset demostrara que los investigadores podrían haber llegado al mismo resultado utilizando 50 pares de gemelos, lanzando una moneda para determinar cuál de cada par estaría en el grupo que consumiría la leche (y ateniéndose a los resultados de ese sorteo). Desde luego, el cálculo estadístico que se utilizaría sería la prueba *t*, tal como la aprendimos en este capítulo, es decir, la prueba *t* para medias dependientes.

Más recientemente, el desarrollo del análisis de la potencia, que presentamos en el capítulo 8, ha reivindicado completamente a Gosset. Ya no quedan dudas de que pueden utilizarse, precisamente, cantidades sorprendentemente pequeñas de participantes cuando el investigador puede encontrar la manera de realizar un diseño de medidas repetidas en el que las diferencias son la unidad básica de análisis (en este caso, cada par de gemelos sería un "participante"). Tal como el mismo Gosset podría haberles dicho, los estudios que utilizan la prueba *t* para medias dependientes pueden tener una sensibilidad extremadamente alta.

Referencias: Peters (1987); Tankard (1984).

CONTROVERSIAS Y LIMITACIONES

Las principales controversias con respecto a la prueba *t* están relacionadas con sus ventajas y desventajas relativas en comparación con varias alternativas, las cuales se discutirán en mayor detalle en le capítulo 15. (Los mismos temas surgen también con respecto a los procedimientos que trataremos en los capítulos 10 al 13). Existe, sin embargo, una consideración que queremos comentar aquí. Esta se relaciona con todos los diseños de investigación en los cuales los mismos participantes se prueban antes y después de alguna intervención experimental. (Es el tipo de situación para la evaluación en la que con frecuencia se utiliza la prueba *t* para medias dependientes).

Medir simplemente a un grupo de personas antes y después de algún procedimiento experimental, sin ningún tipo de grupo control que no experimente el procedimiento, puede tener una potencia alta, pero es un diseño de investigación débil en cuanto a la claridad de las conclusiones que puede producir (Cook & Campbell, 1979). Como se describe detalladamente en el apéndice A, aun cuando tal estudio produzca una diferencia significativa, quedan muchas explicaciones alternativas posibles en cuanto a la razón por la cual ocurrió tal diferencia. Por ejemplo, los participantes podrían haber madurado o mejorado de todos modos durante ese período, o tal vez otros hechos ocurrieron en el transcurso del tiempo entre una prueba y otra, o los participantes que no recibieron beneficios pueden haber abandonado el experimento. Incluso es posible que la propia prueba inicial causara cambios que, de otro modo, no podrían haber ocurrido.

No obstante, es importante observar que las dificultades que presentan las investigaciones en las que se prueba a las personas antes y después de alguna intervención, se comparten sólo levemente con el tipo de estudio en el que los participantes son probados en dos condiciones diferentes, como por ejemplo de ruido y silencio, probando primero a una mitad en unas condiciones y a la otra mitad, también primero, en las otras condiciones.

LAS PRUEBAS *t* SEGÚN SE DESCRIBEN EN LAS PUBLICACIONES CIENTÍFICAS

Las publicaciones científicas describen usualmente las pruebas *t* en un formato bastante estándar que indica los grados de libertad, el punto *t* y el nivel de significación. Por ejemplo, "$t(24) = 2{,}80$, $p < 0{,}05$" indica que el investigador utilizó una prueba *t* con 24 grados de libertad, obtuvo un punto *t* de 2,80, y el resultado fue significativo al nivel 0,05. También puede establecerse si se utilizó una prueba de una o dos colas (si no se indica nada al respecto, debemos suponer que el investigador utilizó una prueba de dos colas). En líneas generales se indican las medias, y a veces los desvíos estándar de cada prueba. Rara vez se indica el desvío estándar de las diferencias.

Si el estudiante del ejemplo acerca de los alumnos del edificio de dormitorios hubiera informado los resultados en una publicación científica, lo hubiera hecho más o menos así: "La muestra tomada del edificio de dormitorios en el que resido produjo una media de 3,2 horas de estudio ($SD = 0{,}80$). Sobre la base de una prueba *t* para una sola muestra (una cola), el resultado era significativamente diferente a la media conocida de 2,5 horas correspondiente a la facultad en general, $t(15) = 3{,}50$, $p < 0{,}01$". Los investigadores del ejemplo ficticio acerca de las víctimas de la inundación podrían haber redactado sus resultados de la siguiente manera: "El grado de esperanza informado por nuestra muestra de víctimas de la inundación ($M = 4{,}7$, $SD = 1{,}89$) no fue significativamente diferente del punto medio de la escala (4,0), $t(9) = 1{,}17$".

Como ya observamos, los psicólogos rara vez utilizan una prueba *t* para una sola muestra. Presentamos esta prueba *t* principalmente como paso previo para la más ampliamente utilizada prueba *t* para muestras dependientes. No obstante, a veces las publicaciones científicas pueden llegar a informar acerca de una prueba t para una sola muestra. Por ejemplo, Weller y Weller (1997) realizaron un estudio acerca de la tendencia de las mujeres que viven juntas a sincronizar sus ciclos menstruales. Para realizar el análisis estadístico, compararon los valores obtenidos por las mujeres que participaron del estudio, de una medida de sincronización de pares de mujeres que viven juntas (población 1), con el grado de sincronización de esos pares de mujeres, esperado en forma casual (población 2). Es decir, crearon una especie de población artificial con una media de lo que se esperaría si no hubiera sincronización, y analizaron los resultados con "pruebas t para una muestra" (p. 147). La tabla 9-11 indica esos resultados. Cada línea de la tabla es una prueba t independiente para una sola muestra. La primera línea es una prueba que compara los registros de sincronización de 6,32 de los 30 pares de hermanas compañeras de cuarto (la muestra de lo que llamaríamos población 1) con un registro de sincronización esperado de 7,76 (lo que denominaríamos la media de la población 2). La línea muestra esos datos más la diferencia de 1,44, el desvío estándar de esta diferencia –que es igual a 3,40–, el punto t de 2,27 y el nivel p de 0,011. Un detalle importante es que la columna t en realidad está escrita como "$t(1)$". Esto no es lo estándar y realmente no significa que su distribución t tenía un grado de libertad. Suponemos que significa que se trata de una prueba t para una sola muestra.

Como mencionamos anteriormente, la prueba t para medias dependientes es mucho más común. Olthoff (1989) podría haber informado del siguiente modo su resultado en el ejemplo que utilizamos anteriormente: "Existía una disminución significativa de la calidad de comunicación, decreciendo de 116,32 antes del matrimonio a 104,26 después del matrimonio, $t(18) = 2{,}76$, $p < 0{,}05$, dos colas". El investigador que realizó el estudio ficticio acerca de los cirujanos podría haber redactado lo siguiente: "La media de desempeño del grupo que trabajó en condición silenciosa fue 18,89, mientras que el rendimiento del grupo que trabajó en condición ruidosa fue 16,89. La diferencia no resultó estadísticamente significativa a nivel 0,01, incluso con una prueba de una cola, $t(8) = 2{,}20$". Para dar otro ejemplo, Holden et. al. (1997) compararon las actitudes informadas por madres con respecto al castigo corporal de sus hijos desde antes hasta después de 3 años de tener a su primer hijo. "El cambio promedio en las actitudes anteriores y actuales de las mujeres fue significativo, $t(107) = 10{,}32$, $p < 0{,}001$..." (p. 485). (El cambio implicó que después de tener a su primer hijo tenían sentimientos más negativos con respecto al castigo corporal).

Los investigadores también presentan con frecuencia las medias de los grupos en una tabla. Por ejemplo, Pezdek y sus colegas (1997) recordaron a cada uno de los integrantes de un grupo de alumnos universitarios varios hechos que supuestamente les habían sucedido cuando tenían ocho años de edad. Se les pidió a los alumnos que describieran el hecho con algún grado de detalle. Estas descripciones fueron clasificadas por cantidad de palabras recordadas y cantidad de unidades de ideas recordadas. También se pidió a los alumnos que clasificaran cada hecho en cuanto a la claridad con la que lo recordaban y en cuanto al nivel de seguridad que tenían con respecto a que el hecho efectivamente había ocurrido. Algunos de los hechos realmente habían ocurrido y algunos podrían haber ocurrido pero no ocurrieron. (Los investigadores se habían comunicado con las madres de los alumnos con anterioridad, con el permiso de los estudiantes). Como es típico en tales investigaciones, muchos de los alumnos equivocadamente recordaron haber experimentado los hechos falsos. Estos son los resultados:

> Para investigar las diferencias potenciales entre recuerdos de hechos reales y recuerdos de hechos falsos, comparamos varias características de los recuerdos de 13 sujetos que recordaron al menos un hecho falso. Se realizaron pruebas de significación de dos colas con esta información, y el resultado aparece en la [tabla 9-12]. En comparación con los recuerdos de hechos falsos, al recordar hechos verdaderos se emplean significativamente más palabras, $t(12) = 4{,}54$, $p < 0{,}001$, y más unidades de ideas, $t(12) = 3{,}43$, $p < 0{,}01$. Por lo tanto, el resultado de los

recuerdos de hechos verdaderos, en comparación con los de hechos falsos, podría ser diferenciado en cuanto a la cantidad de nuevos detalles dados con respecto a cada uno; se dieron casi el doble de detalles con respecto a los hechos verdaderos que a los falsos. En comparación con los recuerdos de hechos falsos, el recuerdo de los hechos verdaderos también estaba relacionado con clasificaciones significativamente superiores de claridad, $t(12) = 3{,}99$, $p < 0{,}01$; y de certeza, $t(12) = 2{,}73$, $p < 0{,}02$ (p. 438).

Es importante observar que en este ejemplo nunca hicieron referencia al nombre de la prueba de significación. Sin embargo, sabemos que se trata de una prueba *t* porque utilizan la *t* al describir los resultados. Además, podemos darnos cuenta de que es una prueba *t* para medias dependientes porque están comparando los valores de cada participante en cuanto al recuerdo de hechos verdaderos y al recuerdo de hechos falsos, cada uno con su propio valor.

Tabla 9-11.
Sincronización menstrual y valores esperados (por días).

Grupo/mes	*N*	Valor de sincronización	Valor esperado	Diferencia	*SD*	*t*(1)	*p*
Compañeras - hermanas							
Mes 1	30	6,32	7,76	1,44	3,40	2,27	0,011
Mes 2	30	6,24	7,76	1,52	3,08	2,66	0,004
Mes 3	29	7,40	7,76	0,36	3,08	0,57	0,280
Amigas íntimas - compañeras							
Mes 1	39	5,73	7,75	2,02	3,84	3,25	<0,000
Mes 2	39	6,01	7,75	1,74	4,25	2,52	0,006
Mes 3	31	7,44	7,75	0,31	4,61	0,88	0,19
Familias							
Mes 1	18	5,80	7,70	1,90	2,74	2,86	<0,000
Mes 2	18	6,09	7,70	1,61	1,89	3,52	<0,000
Mes 3	17	7,19	7,70	0,51	2,71	0,75	0,23

Fuente: Weller, A. & Weller, L. (1997), tab. 1. "Sincronización menstrual en condiciones óptimas: Familias nómades". *Revista Científica de Psicología Comparativa [Journal of Comparative Psychology]*, 111, 143-151. Copyright, 1997, por la Asociación Americana de Psicología [American Psychological Association]. Reimpreso con autorización.

Tabla 9-12.
Medias (y desvíos estándar) de medidas que comparan el recuerdo de hechos verdaderos y falsos del experimento 1.

	Hecho recordado	
Medida	Verdadero	Falso
Cantidad de palabras recordadas ***	27,79 (8,81)	15,42 (7,69)
Cantidad de unidades de ideas recordadas**	6,33 (2,53)	3,23 (1,55)
Puntuación en claridad***	6,90 (0,17)	4,00 (0,18)
Puntuación en certeza**	6,88 (0,21)	5,00 (0,21)

[a] La escala de puntuación iba de 1 (bajo) a 10 (alto).
* $p < 0{,}02$, dos colas; ** $p < 0{,}01$, dos colas; *** $p < 0{,}001$, dos colas.

Fuente: Pezdek, K., Finger, K., & Hodge, D. (1997), tab. 2. "Fijación de falsos recuerdos de la niñez: el papel de la plausibilidad de un evento". *Ciencia Psicológica [Psychological Science]*, 8, 439. Copyright,1997, por la Asociación Americana de Psicología [American Psychological Society]. Reimpreso con autorización.

RESUMEN

Cuando no se conoce la varianza poblacional se utilizan los cinco pasos estándar de la prueba de hipótesis. No obstante, en este caso debemos estimar la varianza poblacional a partir de los valores muestrales, utilizando una fórmula que divide la suma de los desvíos cuadráticos por los grados de libertad ($gl = N - 1$). Además, cuando no se conoce la varianza, la distribución comparativa de medias es una distribución t (cuyos puntos de corte se indican en una tabla t). Una distribución t posee colas ligeramente más pesadas que las de una curva normal (exactamente en qué medida son más pesadas depende de cuán pocos sean los grados de libertad). Finalmente, en este caso, se denomina punto t a la cantidad de desvíos estándar con respecto a la media a la que se encuentra la media muestral en la distribución t.

La prueba t para medias dependientes se utiliza en estudios en los que cada participante presenta dos valores, como por ejemplo un valor anterior y uno posterior. En esta prueba t, primero se calcula una diferencia para cada participante; luego se realizan los usuales cinco pasos de la prueba de hipótesis con las modificaciones descriptas en el párrafo anterior y se convierte a la población 2 en una población de diferencias con una media de 0 (ausencia de diferencia).

Un supuesto de la prueba t es que la distribución poblacional es una curva normal. Sin embargo, aun cuando no lo sea, la prueba t usualmente es bastante exacta. La principal excepción en el caso de la prueba t para medias dependientes es cuando la población de diferencias es altamente asimétrica y trabajamos con una prueba de una cola.

El tamaño del efecto de un estudio en el que se utiliza una prueba t para medias dependientes es la media de las diferencias dividida por el desvío estándar de esas diferencias. Existen tablas especiales en las que se pueden encontrar la potencia y el tamaño de muestra necesarios para obtener una potencia del 80%. La potencia de estudios en los que se utilizan diferencias es usualmente mucho mayor que el de aquellos estudios en los que se utilizan otros diseños con la misma *cantidad de participantes.*

Investigadores expertos en metodología señalan que las investigaciones que involucran a un sólo grupo probado antes y después de algún hecho interpuesto, sin un grupo control, deja abiertas muchas explicaciones alternativas de cualquier cambio observado.

En las publicaciones científicas, las pruebas t se informan utilizando un formato estándar, por ejemplo, "$t(24) = 2{,}80$, $p < 0{,}05$".

Términos clave

- Supuesto.
- Estimación sesgada.
- Grados de libertad (gl).
- Diferencias.
- Diseño de medidas repetidas.
- Robustez.
- Distribución t.
- Punto t.
- Tabla t.
- Pruebas t
- Prueba t para una sola muestra.
- Prueba t para medias dependientes.
- Estimación no sesgada de la varianza poblacional (S^2).

Ejercicios

Los ejercicios implican la realización de cálculos (con la ayuda de una calculadora). La mayoría de los problemas estadísticos reales se resuelven por computadora, pero aunque exista la posibilidad de utilizarla, es conveniente realizar estos ejercicios manualmente para incorporar el método de trabajo.

Para adquirir práctica en la utilización de una computadora, para resolver problemas estadísticos, se puede utilizar la sección de computación de cada capítulo, publicada en la *Guía de estudio y libro de tareas de computación para el alumno [Student's Study Guide and Computer Workbook]* que acompaña este libro.

Todos los datos de esta sección son ficticios (a menos que se especifique lo contrario).

Las respuestas a los ejercicios de la serie I se encuentran al final del libro.

SERIE I

1. En cada uno de los estudios que aparecen a continuación, se está comparando la media de una sola muestra con una población de la cual se conoce la media pero no la varianza. Decida si el resultado de cada uno de estos estudios es o no significativo.

	Tamaño muestral (N)	Media poblacional (μ)	Varianza poblacional estimada (S^2)	Media muestral (M)	Colas	Nivel de signific. (α)
(a)	64	12,40	9,00	11,00	1 (predicción baja)	0,05
(b)	49	1.006,35	317,91	1.009,72	2	0,01
(c)	400	52,00	7,02	52,41	1 (predicción alta)	0,01

2. Supongamos que un candidato que se postula como jefe de policía afirma que reducirá el tiempo promedio de respuesta a emergencias a menos de 30 minutos, que es considerado el tiempo de respuesta promedio para emergencias bajo el mandato del jefe de policía actual. No existen registros anteriores, por lo tanto, no podemos determinar el desvío estándar real de esos tiempos de respuesta. Gracias a esta campaña, él es elegido jefe de policía, y ahora se guardan los registros cuidadosamente. Los tiempos de respuesta durante el primer mes son 26, 30, 28, 29, 25, 28, 32, 35, 24 y 23 minutos.

Utilizando un nivel de significación del 5%, ¿cumplió él su promesa? a) Realice los cinco pasos de la prueba de hipótesis; b) ilustre su respuesta con un histograma de la distribución muestral y gráficos de la distribución poblacional y la distribución de medias, indique el punto t y los puntos de corte correspondientes al nivel de significación seleccionado, y c) explique su respuesta a alguien que nunca ha tomado un curso de estadística.

3. Para cada uno de los siguientes estudios en los que se utilizan diferencias, determine si la diferencia media es significativamente diferente de 0. Además, calcule el tamaño del efecto (si en la tabla no se indican los gl, utilice el t correspondiente al valor gl menor más cercano).

	Cantidad de diferencias de la muestra	Media de las diferencias de la muestra	Varianza poblacional est. de las dif.	Colas	Nivel de signific.
(a)	20	1,7	8,29	1 (predicción alta)	0,05
(b)	164	2,3	414,53	2	0,05
(c)	15	-2,2	4,00	1 (predicción baja)	0,01

4. En cuatro ciudades del Valle Central de California se implementó, en agosto de 1997, un programa para reducir la cantidad de desperdicios. La cantidad de basura en las calles (cantidad promedio en libras de basura recolectada por manzana, por día) se midió durante el mes de julio anterior al comienzo del programa y, luego, el siguiente julio, después de que el programa hubiera estado en efecto durante un año. Los resultados fueron los siguientes:

Ciudad	Julio 1997	Julio 1998
Fresno	19	2
Merced	10	4
Bakersfield	18	9
Stockton	19	1

Utilizando un nivel de significación del 1%, ¿hubo una disminución significativa de la cantidad de desperdicios? a) Realice los cinco pasos de la prueba de hipótesis; b) ilustre su respuesta con un histograma de la distribución muestral y con gráficos de la distribución poblacional y la distribución de medias, indique el punto t y los puntos de corte correspondientes al nivel de significación seleccionado; c) calcule el tamaño del efecto, y d) explique su respuesta a alguien que comprende los conceptos de *media*, *desvío estándar* y *varianza* pero que no sabe nada más sobre estadística.

5. ¿Cuál es la potencia de cada uno de los siguientes estudios (sobre la base de un nivel de significación de 0,05)?

	Tamaño del efecto	N	Colas
(a)	Pequeño	20	1
(b)	Mediano	20	1
(c)	Mediano	30	1
(d)	Mediano	30	2
(e)	Grande	30	2

6. Un psicólogo realiza un estudio acerca de ilusiones perceptivas en dos condiciones diferentes de iluminación. Veinte participantes fueron probados cada uno en las dos condiciones distintas. El experimentador informó: "La media de ilusiones efectivas fue 6,72 en condiciones de luminosidad y 6,85 en condiciones de iluminación débil, una diferencia no significativa, $t(19) = 1,62$". Explique el resultado a una persona que nunca ha asistido a un curso sobre estadística. Asegúrese de utilizar en su respuesta gráficos de las distribuciones.

7. Se realizó un estudio acerca de las características de la personalidad a 100 alumnos que fueron probados al comienzo y al final de su primer año de facultad. Los investigadores informaron los resultados en la siguiente tabla:

Escala de personalidad	Otoño		Primavera		Diferencia	
	M	SD	M	SD	M	SD
Angustia	16,82	4,21	15,32	3,84	1,50**	1,85
Depresión	89,32	8,39	86,24	8,91	3,08**	4,23
Introversión	59,89	6,87	60,12	7,11	0,23	2,22
Neurosis	38,11	5,39	37,32	6,02	0,89*	4,21

$^*p < 0,05$; $^{**}p < 0,01$.

a) Concentrándose en las diferencias, calcule los valores t para cada escala de personalidad. (Considere que los SD de la tabla corresponden a lo que hemos clasificado como S, la estimación no sesgada del desvío estándar de la población). b) Explique el significado de la tabla a una persona que nunca ha asistido a un curso de estadística.

SERIE II

1. En cada uno de los siguientes estudios, se compara a la media de una sola muestra con una población de la cual se conoce la media pero no la varianza. Decida, en cada caso, si el resultado es o no significativo. (Si los *gl* no aparecen en la tabla, utilice el t correspondiente al valor *gl* menor más cercano). Asegúrese de indicar todos sus cálculos.

	Tamaño muestral (N)	Media poblacional (μ)	Desvío estándar estimado (S)	Media muestral (M)	Colas	Nivel de signific. (α)
(a)	16	100,31	2,00	100,98	1 (predicción alta)	0,05
(b)	16	0,47	4,00	0,00	2	0,05
(c)	16	68,90	9,00	34,00	1 (predicción baja)	0,01

2. Existen teorías biológicas que sostienen que los humanos se han adaptado a su ambiente físico. Una de estas teorías sostiene la hipótesis de que las personas seguirían espontaneamente un ciclo de 24 horas de sueño y vigilia, aun cuando no fueran expuestas al patrón usual de la luz solar. Para probar esta noción, ocho voluntarios contratados fueron ubicados (individualmente) en una habitación en la que no había luz del exterior, ni relojes, ni ninguna otra indicación del transcurso del tiempo. Podían encender o apagar las luces cuando quisieran. Después de un mes en la habitación, cada individuo mostró una tendencia a desarrollar un ciclo estable. Sus ciclos al finalizar el estudio fueron los siguientes: 25, 27, 25, 23, 24, 25, 26 y 25.

Utilizando un nivel de significación del 5%, ¿qué conclusión sacaríamos con respecto a la teoría de que 24 horas es el ciclo natural? (Es decir, la duración promedio del ciclo en estas condiciones es significativamente diferente al de 24 horas?). a) Realice los cinco pasos de la prueba de hipótesis. b) Ilustre su respuesta con un histograma de la distribución muestral y gráficos de la distribución poblacional y la distribución de medias, e indique el punto *t* y los puntos de corte correspondientes al nivel de significación seleccionado. c) Explique su respuesta a alguien que nunca ha asistido a un curso de estadística.

3. Cuatro individuos con alto nivel de colesterol iniciaron una dieta intensiva: evitan las comidas con alto contenido de colesterol y toman suplementos especiales. Sus niveles de colesterol antes y después de la dieta fueron los siguientes:

Participante	Antes	Después
J.K.	287	255
L.M.M	305	269
A.K.	243	245
R.O.S.	309	247

Utilizando un nivel de significación del 5%, ¿se produjo un cambio significativo del nivel de colesterol? a) Realice los cinco pasos de la prueba de hipótesis. b) Ilustre su respuesta con un histograma de la distribución muestral y con gráficos de la distribución poblacional y la distribución de medias, e indique el punto *t* y los puntos de corte correspondiente al nivel de significación seleccionado. c) Calcule el tamaño del efecto. d) Explique su respuesta a alguien que nunca ha asistido a un curso de estadística.

4. Un tribunal ordenó a cinco personas penadas por exceso de velocidad a que asistieran a un taller. Un mecanismo especial incorporado en sus autos mantuvo un registro de sus velocidades durante 2 semanas antes y después de participar del taller. Las velocidades máximas de cada persona durante 2 semanas antes y 2 semanas después de participar del taller fueron las siguientes:

Participante	Antes	Después
L.B.	65	58
J.K.	62	65
R.C.	60	56
R.T.	70	66
J.M.	68	60

Utilizando un nivel de significación del 5%, ¿deberíamos concluir que es probable que una persona conduzca a menor velocidad después de participar de un taller de trabajo? a) Realice los cinco pasos de la prueba de hipótesis. b) Ilustre su respuesta con un histograma de la distribución muestral y con gráficos de la distribución poblacional y la distribución de medias, e indique el punto *t* y los puntos de corte correspondiente al nivel de significación seleccionado. c) Calcule el tamaño del efecto. d) Explique su respuesta a alguien que está familiarizado con la prueba de hipótesis con poblaciones conocidas, pero que nunca ha aprendido nada sobre las pruebas *t*.

5. Se midió la cantidad de oxígeno consumido por seis individuos durante dos periodos de 10 minutos mientras permanecían sentados con los ojos cerrados. Durante un periodo, escuchaban una excitante historia de aventuras; durante el otro, escuchaban música tranquila. (El orden de las condiciones era el opuesto para una mitad de los participantes).

Participante	Historia	Música
1	6,12	5,39
2	7,25	6,72
3	5,70	5,42
4	6,40	6,16
5	5,82	5,96
6	6,24	6,08

Sobre la base de los resultados indicados ¿es menor el consumo de oxígeno cuando escuchan música? Utilice un nivel de significación del 1%. a) Realice los cinco pasos de prueba de hipótesis. b) Ilustre su respuesta con un histograma de la distribución muestral y con gráficos de la distribución poblacional y de la distribución de medias, e indique el punto *t* y los puntos de corte correspondiente al nivel de significación seleccionado. c) Calcule el tamaño del efecto. d) Explique su respuesta a alguien que comprende el concepto de media, desvío estándar y varianza pero que no sabe nada más sobre estadística.

6. A cinco alumnos de segundo año se les tomó un examen de evaluación del nivel de inglés antes y después de recibir instrucciones sobre gramática básica. Sus registros fueron los siguientes:

Estudiante	Antes	Después
A	20	18
B	18	22
C	17	15
D	16	17
E	12	9

¿Es razonable concluir que futuros alumnos lograrían registros más altos después de recibir las instrucciones? Utilice un nivel de significación del 5%. a) Realice los cinco pasos de la prueba de hipótesis. b) Ilustre su respuesta con un histograma de la distribución de muestras y con gráficos de la distribución poblacional y de la distribución de medias, e indique el punto t y los puntos de corte de significación. c) Calcule la magnitud de efecto. d) Explique su respuesta a alguien que comprende el concepto de media, desvío estándar y varianza pero que no sabe nada más sobre estadística.

7. Se realizó un estudio comparando la actividad sindical de empleados de 10 plantas durante dos décadas diferentes. El investigador informó "un aumento significativo de la actividad sindical, $t\ (9) = 3{,}28$, $p < 0{,}01$". Explique este resultado a una persona que nunca ha tomado un curso de estadística. Asegúrese de utilizar gráficos de las distribuciones en su respuesta.

Apéndice del capítulo: fórmulas de cálculo opcionales para la prueba *t* para medias dependientes

Para realizar una prueba t para medias dependientes, después de convertir los valores en diferencias, los pasos usuales son: a) calcular la media de las diferencias; b) calcular la suma de los desvíos cuadráticos de las diferencias con respecto a la media de esas diferencias; c) calcular la varianza estimada de la distribución poblacional de diferencias individuales (S^2); d) calcular el desvío estándar estimado de la distribución de medias de las diferencias (S_M), y e) calcular el punto t. Combinando algunos de estos pasos y aplicando algunas manipulaciones algebraicas, una vez que hemos convertido todos los datos en diferencias, podemos utilizar las siguientes fórmulas de cálculo para encontrar S y t.

$$S = \sqrt{\frac{\Sigma X^2 - \frac{(\Sigma X)^2}{N}}{N-1}} \tag{9-10}$$

$$t = \frac{\Sigma \frac{X}{N}}{\frac{S}{\sqrt{N}}} \tag{9-11}$$

La tabla 9-13 indica el cálculo de la prueba *t* para medias dependientes correspondiente al estudio ficticio acerca de la coordinación entre el pulso y la vista de cirujanos, utilizando las fórmulas de cálculo. Compare estos cálculos con los de la tabla 9-6 correspondiente a la misma información pero utilizando fórmulas de definición.

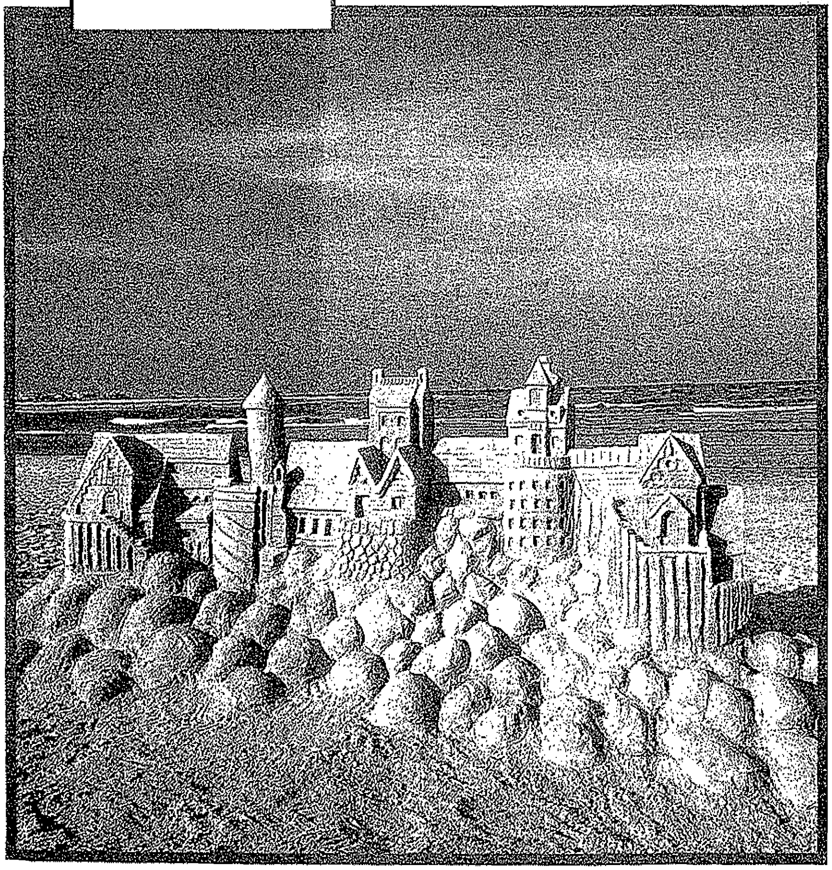
Tabla 9-13.
Cálculo de la prueba *t* para medias dependientes correspondiente al ejemplo acerca de la coordinación entre el pulso y la vista de cirujanos, utilizando las fórmulas de cálculo. (Datos ficticios).

Cirujano	**Condiciones**		**Diferencia**	**Diferencia cuadrática**
	Silencio	Ruido	(X)	(X^2)
1	18	12	6	36
2	21	21	0	0
3	19	16	3	9
4	21	16	5	25
5	17	19	−2	4
6	20	19	1	1
7	18	16	2	4
8	16	17	−1	1
9	20	16	4	16
Σ:	170	152	18	96

$$S = \sqrt{\frac{\Sigma X^2 - (\Sigma X)^2/N}{N-1}} = \sqrt{\frac{96 - 18^2/9}{9-1}} = \sqrt{\frac{96 - 324/9}{8}} = \sqrt{\frac{96-36}{8}} = \sqrt{\frac{60}{8}} = \sqrt{7{,}5} = 2{,}74$$

$$t = \frac{\Sigma X/N}{S/\sqrt{N}} = \frac{18/9}{2{,}74/\sqrt{9}} = \frac{2}{2{,}74/3} = \frac{2}{0{,}91} = 2{,}20$$

10 Prueba t para medias independientes



Descripción del capítulo

- ▶ Estrategia básica de la prueba *t* para medias dependientes: la distribución de diferencias entre medias.
- ▶ Pasos de la prueba de hipótesis con una prueba *t* para medias independientes.
- ▶ Premisas de la prueba *t* para medias independientes.
- ▶ Tamaño del efecto y potencia de la prueba *t* para medias independientes.
- ▶ Controversias y limitaciones.
- ▶ La prueba *t* para medias independientes según se describen en los artículos de investigación.
- ▶ Resumen.
- ▶ Términos clave.
- ▶ Problemas prácticos.
- ▶ Apéndice del capítulo: fórmulas de cálculo opcionales de la prueba *t* para medias independientes.

Este capítulo analiza la prueba de hipótesis para los casos en los que se comparan dos muestras, tales como un grupo experimental y un grupo de control. Son situaciones en las que se realiza una prueba *t* debido a que las varianzas poblacionales no se conocen y, por lo tanto, deben estimarse. En este caso, la prueba se denomina **prueba *t* para medias independientes,** porque se comparan medias de dos grupos de personas completamente separados, cuyos valores son independientes el uno del otro. La prueba *t* para medias independientes se contrapone con la prueba *t* para medias dependientes analizada en el capítulo anterior, en la que había dos grupos de valores, pero ambos provenían del mismo grupo de personas (como es el caso de las mismas personas medidas antes y después de un programa de asesoramiento).

ESTRATEGIA BÁSICA DE LA PRUEBA *t* PARA MEDIAS INDEPENDIENTES: LA DISTRIBUCIÓN DE DIFERENCIAS ENTRE MEDIAS

La prueba *t* para medias independientes funciona de la misma manera que la prueba de hipótesis que ya hemos aprendido, con una excepción fundamental: el resultado clave del estudio es una diferencia entre las medias de las dos muestras. Por lo tanto, la distribución comparativa debe ser una **distribución de diferencias entre medias.**

Contenido de una distribución de diferencias entre medias

La distribución especial a la que nos referimos se encuentra, en un sentido, a dos pasos de las poblaciones de observaciones de individuos: en primer lugar, tenemos una distribución de medias por cada población de observaciones de individuos, y luego construimos una distribución de diferencias entre pares de medias (cada una de las cuales proviene de una de esas dos distribuciones de medias). Pensemos que la distribución de diferencias entre medias se construye de la siguiente forma: a) se selecciona al azar una media de la distribución de medias de la población 1, b) se selecciona al azar una media de la distribución de medias de la población 2 y c) se resta (es decir tomamos la media de la distribución de medias de la población 1 y le restamos la media de la distribución de medias de la población 2). El resultado es una diferencia entre las dos medias seleccionadas. Luego se repite el proceso creando una segunda diferencia, es decir, la diferencia entre las nuevas medias seleccionadas. Repitiendo este proceso una gran cantidad de veces se crea una distribución de diferencias entre medias.

Ilustración de la lógica general de la prueba *t* para medias independientes

La figura 10-1 representa gráficamente la creación lógica completa que implica una distribución de diferencias entre medias. En la parte superior están las dos distribuciones poblacionales cuyas características desconocemos. No obstante, sí sabemos que si la hipótesis nula es verdadera, las dos medias poblacionales son iguales, es decir, la hipótesis nula establece que $\mu_1 = \mu_2$. También podemos estimar las varianzas poblacionales sobre la base de la información obtenida a través de la muestra (las varianzas estimadas serán S_1^2 y S_2^2).

Debajo de cada distribución poblacional se encuentra la distribución de medias correspondiente a esa población. Utilizando la varianza poblacional estimada y sabiendo el tamaño de cada muestra, podemos calcular la varianza de cada distribución de medias utilizando el método acostumbrado (es decir, la varianza poblacional de origen dividida por el tamaño de la muestra).

Debajo de las dos distribuciones de medias, y creada a partir de ellas, se encuentra la crucial distribución de diferencias entre medias. Dado que la varianza de esta distribución se estima finalmente sobre la base de las varianzas poblacionales estimadas, podemos considerarla una distribución *t*. El objetivo de una prueba *t* para medias independientes es decidir si la diferencia entre las medias de las dos muestras reales es más extrema que la diferencia de corte en la distribución de diferencias. Las dos muestras reales aparecen (como histogramas) en la parte inferior.

No debemos olvidar que todo el procedimiento es, en realidad, una especie de complicado castillo en el aire. Existe sólo en nuestras mentes para ayudarnos a tomar una decisión basada en los resultados de un experimento real. La única realidad concreta en todo este procedimiento son

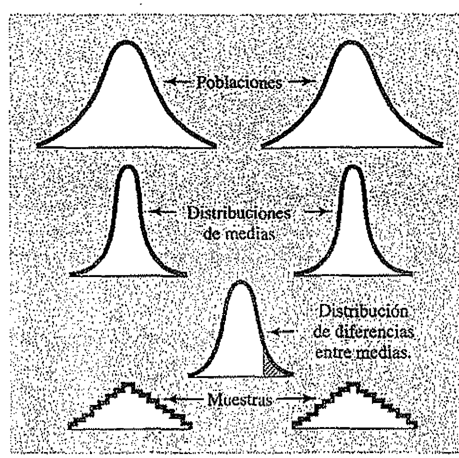


Figura 10-1. Pasos para la creación de una distribución de diferencias de medias.

las dos muestras de valores realmente observadas. Las varianzas poblacionales se estiman sobre la *base de esos valores muestrales. Las varianzas de las dos distribuciones de medias se basan completamente* en las varianzas poblacionales estimadas (y en los tamaños de las muestras). Y, como veremos pronto, las características de la distribución de diferencias entre medias se basan en las dos distribuciones de medias que mencionamos anteriormente.

Aun así, el procedimiento es poderoso. Tiene el poder de la matemática y una lógica implícita: ayuda a desarrollar un conocimiento general basado en los datos específicos de un estudio en particular.

Teniendo una visión general de la lógica básica, ahora nos dedicaremos a cinco detalles clave: a) la media de la distribución de diferencias entre medias, b) la varianza poblacional estimada, c) la varianza y el desvío estándar de la distribución de diferencias entre medias, d) la forma de la distribución de diferencias entre medias y e) el punto *t* correspondiente a la diferencia entre las dos medias particulares que están siendo comparadas.

Media de la distribución de diferencias de medias

En una prueba *t* para medias independientes se tienen en cuenta dos poblaciones; por ejemplo, una población de la cual se extrae el grupo experimental y otra población de la cual se extrae el grupo de control. En la práctica, el investigador no conoce la media de ninguna de las poblaciones, pero efectivamente sabe que si la hipótesis nula es verdadera, esas dos poblaciones tienen la misma media. Si las dos poblaciones tienen la misma media, la distribución de medias de cada una de ellas tendrá también la misma media. Si se seleccionan al azar dos muestras de dos distribuciones con la misma media, las diferencias de las medias de estas muestras aleatorias, a la larga, deberían compensarse en 0. El resultado de toda esta lógica es que cualesquiera sean los datos específicos del estudio, el investigador sabe que si la hipótesis nula es verdadera, la distribución de diferencias de medias tiene una media de 0.

Estimación de la varianza poblacional

En el capítulo 9 aprendimos a estimar la varianza poblacional utilizando los valores muestrales. Era el resultado de la suma de los desvíos cuadráticos dividido por los grados de libertad (el tamaño de la muestra menos 1).

Para realizar una prueba *t* para medias independientes, debería ser lógico suponer que las poblaciones de las cuales provienen las dos muestras tienen la misma varianza. (Si la hipótesis nula es verdadera, también tienen la misma media. Pero sea la hipótesis nula verdadera o no, debemos estar en condiciones de suponer que las dos poblaciones tienen la misma varianza). Por lo tanto, cuando estimamos la varianza a partir de los valores de cada muestra estamos obteniendo dos estimaciones separadas de lo que debería ser el mismo número (la varianza de cada muestra es una estimación de lo que se supone que es igual para las dos poblaciones). En la práctica, las dos estimaciones no serán idénticas, pero como se supone que las dos están estimando lo mismo, la mejor solución es promediar las dos estimaciones para obtener la mejor estimación, única y general. A esto se lo denomina **estimación combinada de la varianza poblacional** ($S^2_{\text{Combinada}}$).

Al realizar el promedio, también debemos tener en cuenta el hecho de que si una muestra es mayor que la otra, es probable que la estimación que produzca sea más precisa (porque se basa en mayor información). Si las dos muestras fueran exactamente del mismo tamaño, podríamos simplemente sacar un promedio simple de las dos estimaciones. Pero cuando no lo son, necesitamos realizar ciertos ajustes en nuestro promedio para dar más valor relativo a la muestra mayor. Necesitamos un **promedio ponderado**, un promedio relativizado según la cantidad de información que proporciona cada muestra. Para ser precisos, lo que importa no es la cantidad de valores de cada muestra sino la cantidad de grados de libertad (la cantidad de valores menos 1).

Por lo tanto, cuando creamos un promedio ponderado éste tiene que basarse en los grados de libertad. El procedimiento consiste en calcular en qué proporción contribuye cada muestra con los grados de libertad totales; después multiplicamos esa proporción por la estimación proveniente de cada muestra y, finalmente, sumamos los dos resultados y obtenemos la estimación ponderada. La fórmula que expresa el principio que acabamos de describir es la siguiente:

$$S^2_{\text{Combinada}} = \frac{gl_1}{gl_{\text{Total}}}\left(S_1^2\right) + \frac{gl_2}{gl_{\text{Total}}}\left(S_2^2\right) \qquad (10\text{-}1)$$

En la fórmula precedente, $S^2{}_{\text{Combinada}}$ es la estimación combinada de la varianza poblacional, gl_1 son los grados de libertad correspondientes a la población 1, y gl_2 son los grados de libertad correspondientes a la población 2. (No debemos olvidar que cada gl es la cantidad de valores muestrales menos 1). gl_{Total} son los grados de libertad totales ($gl_{\text{Total}} = gl_1 + gl_2$). S_1^2 es la estimación de la varianza poblacional sobre la base de los valores de la muestra que proviene de la población 1; S_2^2 es la estimación sobre la base de los valores de la muestra que proviene de la población 2.

Analicemos un estudio en el que la estimación de la varianza poblacional, sobre la base de un grupo experimental de 11 participantes, es 60, y la estimación de la varianza poblacional sobre la base de un grupo de control de 31 participantes es 80. La estimación del grupo experimental se basa en 10 grados de libertad (11 participantes menos 1); la estimación del grupo de control se basa en 30 grados de libertad (31 participantes menos 1). La información total sobre la que se basa la estimación son los grados totales de libertad, en este caso, 40. Por lo tanto, el grupo experimental proporciona un cuarto de la información (10/40 = 1/4), y el grupo control proporciona tres cuartos de la información (30/40 = 3/4).

Después multiplicamos la estimación del grupo experimental por 1/4 y obtenemos 15 (es decir, 60 x 1/4 = 15), y la estimación del grupo de control por 1/4 y obtenemos 60 (es decir, 80 x 3/4 = 60). Sumando los dos resultados obtenemos una estimación de 15 más 60, es decir de 75. Aplicando la fórmula:

$$S^2_{\text{Combinada}} = \frac{gl_1}{gl_{\text{Total}}}\left(S_1^2\right) + \frac{gl_2}{gl_{\text{Total}}}\left(S_2^2\right) = \frac{10}{40}(60) + \frac{30}{40}(80)$$
$$= \frac{1}{4}(60) + \frac{3}{4}(80) = 15 + 60 = 75$$

Cabe mencionar que este procedimiento no da el mismo resultado que un promedio simple (sin ponderar). Un promedio simple daría una estimación de 70 (es decir, [60+80]/2 = 70). Nuestra estimación combinada ponderada, igual a 75, está más cerca de la estimación realizada sólo sobre la base del grupo de control que de la estimación realizada únicamente sobre la base del grupo experimental. Así es como debe ser, porque la estimación del grupo de control se basó en mayor información. Por otro lado, aún sigue siendo un tipo de promedio. Será evidente que hemos cometido un error en los cálculos si este número no se encuentra entre las dos estimaciones. (También resultará evidente el error de cálculo si no obtenemos un número más cercano a la estimación que proviene de la muestra mayor).

Cálculo de la varianza de cada una de las dos distribuciones de medias

La estimación combinada de la varianza poblacional es la mejor estimación para ambas poblaciones. (No debemos olvidar que para realizar una prueba *t* para medias independientes, debemos estar en condiciones de suponer que las dos poblaciones tienen la misma varianza). Sin embargo, aunque las dos poblaciones tienen la misma varianza, las distribuciones de medias tomadas de ellas usualmente no tienen la misma varianza, ya que la varianza de una distribución de medias es la varianza poblacional dividida por el tamaño de la muestra. Por lo tanto, aun cuando la varianza poblacional sea la misma para las dos poblaciones, si los tamaños de las muestras son diferentes, entonces las dos distribuciones de medias tendrán diferentes varianzas. Expresado en fórmulas:

$$S_{M_1}^2 = \frac{S_{\text{Combinada}}^2}{N_1} \tag{10-2}$$

y

$$S_{M_2}^2 = \frac{S_{\text{Combinada}}^2}{N_2} \tag{10-3}$$

Analicemos nuevamente el ejemplo del estudio en el que había 11 individuos en el grupo experimental y 31 en el grupo de control. En ese ejemplo, descubrimos que la estimación combinada de la varianza poblacional era 75. Por lo tanto, para el grupo experimental, la varianza de la distribución de medias sería 75/11, es decir 6,82; y en el grupo de control, la varianza sería 75/31, es decir, 2,42. (Es importante recordar que al calcular varianzas estimadas dividimos por los grados de libertad, pero cuando calculamos la varianza de una distribución de medias, que no involucra ninguna estimación adicional, dividimos por la cantidad real de observaciones en la muestra). Aplicando las fórmulas,

$$S_{M_1}^2 = \frac{S_{\text{Combinada}}^2}{N_1} = \frac{75}{11} = 6{,}82$$

$$S_{M_2}^2 = \frac{S_{\text{Combinada}}^2}{N_2} = \frac{75}{31} = 2{,}42$$

Varianza y desvío estándar de la distribución de diferencias de medias

La **varianza de la distribución de diferencias de medias** ($S^2_{\text{Diferencia}}$) es la suma de la varianza de la distribución de medias proveniente de la población 1 y la varianza de la distribución de medias proveniente de la población 2. Esto se debe a que, al calcular una diferencia entre dos números, la variación de cada uno contribuye a la variación total de la diferencia. Es como restar un número en movimiento de un objetivo en movimiento. Se representa por la fórmula:

$$S_{\text{Diferencia}}^2 = S_{M_1}^2 + S_{M_2}^2 \tag{10-4}$$

El desvío estándar de la distribución de diferencias de medias ($S_{\text{Diferencia}}$) es la raíz cuadrada de la varianza:

$$S_{\text{Diferencia}} = \sqrt{S_{\text{Diferencia}}^2} \tag{10-5}$$

Analicemos nuevamente el ejemplo del estudio con 11 individuos en el grupo experimental y 31 en el grupo de control. Descubrimos que la varianza de la distribución de medias del grupo experimental era 6,82, y la varianza de la distribución de medias del grupo de control era 2,42. La varianza de la distribución de diferencias entre medias sería entonces 6,82 más 2,42, lo que da un total de 9,24. Por lo tanto, el desvío estándar de esta distribución es la raíz cuadrada de 9,24, que es 3,04. La fórmula es la siguiente:

$$S^2_{\text{Diferencia}} = S^2_{M_1} + S^2_{M_2} = 6{,}82 + 2{,}42 = 9{,}24$$

$$S_{\text{Diferencia}} = \sqrt{S^2_{\text{Diferencia}}} = \sqrt{9{,}24} = 3{,}04$$

Forma de la distribución de diferencias de medias

La distribución de diferencias de medias se basa en la utilización de varianzas poblacionales estimadas; por lo tanto, la distribución comparativa es una distribución t. La estimación de la varianza de esa distribución se basa en estimaciones en las que se utilizan dos muestras; por lo tanto, los grados de libertad de esa distribución t son la suma de los grados de libertad de las dos muestras. Se representa bajo la fórmula,

$$gl_{\text{Total}} = gl_1 + gl_2 \tag{10-6}$$

La novedad en este caso es que gl_{Total}, los grados de libertad totales de ambas muestras juntas, son también los grados de libertad de la distribución t.

En el ejemplo con un grupo experimental de 11 y un grupo de control de 31, los grados de libertad totales serían 40 (es decir, 11 – 1 = 10; 31 – 1 = 30, y 10 + 30 = 40). Para determinar el punto t necesario para la significación, buscamos el punto de corte en la tabla t en la línea correspondiente a los 40 grados de libertad. Supongamos que estuviéramos realizando una prueba de una cola utilizando un nivel de significación de 0,05. La tabla t indica que con 40 grados de libertad, para un resultado significativo, la diferencia entre las medias debe ser de al menos 1,684 desvíos estándar por encima de la diferencia media de 0 en la distribución de diferencias de medias.

El punto t para la diferencia entre las dos medias reales

El punto t que calculamos en el paso 4 de la prueba de hipótesis se encuentra de la siguiente manera: primero, calculamos la diferencia entre las dos medias (es decir, restándole una a la otra). Después, calculamos dónde se ubica esa diferencia en la distribución de diferencias entre medias, es decir, dividimos la diferencia por el desvío estándar de esa distribución. Se expresa bajo la fórmula:

$$t = \frac{M_1 - M_2}{S_{\text{Diferencia}}} \tag{10-7}$$

Por ejemplo, supongamos que la media de la primera muestra es 198 y la media de la segunda muestra es 190. La diferencia entre estas dos medias es 8 (es decir, 198 – 190 = 8). Calculamos que el desvío estándar de la distribución de diferencias entre muestras es, en este ejemplo, 3,04, es decir, un punto t de 2,63 (8/3,04 = 2,63). En otras palabras, en este ejemplo la diferencia de las dos medias se encuentra 2,63 desvíos estándar por encima de la media de la distribución de diferencias de medias. Se expresa bajo la fórmula,

$$t = \frac{M_1 - M_2}{S_{\text{Diferencia}}} = \frac{198-180}{3{,}04} = \frac{8}{3{,}04} = 2{,}63$$

PASOS DE LA PRUEBA DE HIPÓTESIS CON UNA PRUEBA *t* PARA MEDIAS INDEPENDIENTES

Sobre la base de los cinco pasos de la prueba de hipótesis, existen tres aspectos nuevos de una prueba *t* para medias independientes: a) la distribución comparativa ahora es una distribución de diferencias de medias (afecta el paso 2), b) los grados de libertad para encontrar el punto de corte en la tabla *t* se basan en dos muestras (afecta el paso 3) y c) el valor muestral se basa en la diferencia entre las dos medias (afecta el paso 4).

Ejemplo de prueba *t* para medias independientes

Norman y Aron (1997) realizaron una serie de experimentos para probar la predicción de su teoría, la cual planteaba que las parejas que realizan juntas actividades excitantes aumentan su satisfacción marital. En el primero de los estudios, parejas casadas se acercaron al laboratorio para una evaluación. Desde el punto de vista de las parejas, la sesión implicaba llenar cuestionarios acerca de sus matrimonios, ser filmados en video mientras interactuaban en una actividad inusual, y después llenar más cuestionarios acerca de su matrimonio. Sin embargo, sin que las parejas lo supieran, la primera serie de cuestionarios en realidad era un pretexto, la actividad era una manipulación experimental, y la segunda serie de cuestionarios era una prueba posterior. Las actividades se establecieron de tal forma que a algunas parejas les tocaron actividades excitantes y a otras les tocaron actividades neutras. La actividad excitante era un juego cooperativo fisiológicamente excitante; la actividad de control era una tarea física lenta y repetitiva que cada uno de los miembros de la pareja realizaba por su cuenta. De las 28 parejas que participaron en el estudio, 15 se asignaron al azar a la actividad excitante y 13 a la actividad de control. (El plan original había consistido en tener cantidades iguales en los dos grupos, pero algunas parejas que participaron no pudieron ser utilizadas por distintas razones, como por ejemplo, que uno de los miembros de la pareja no llenara correctamente los cuestionarios).

Siguiendo los procedimientos típicos de los experimentos relacionados con la psicología social, los investigadores incluyeron algunas preguntas al final de la sesión para controlar si la manipulación experimental había producido las experiencias esperadas durante la actividad. La técnica se denomina "control de manipulación". Por ejemplo, las preguntas de control de manipulación incluían: "¿En qué medida le resultó excitante la tarea?" y "¿cuán interesante le pareció la tarea?"

La figura 10-2 representa gráficamente la prueba *t* que compara los valores de control de manipulación correspondientes a las dos condiciones; la tabla 10-1 indica los valores y cálculos del mismo ejemplo. Sigamos también los cinco pasos completos de la prueba de hipótesis.

1. Replantear el problema en función de hipótesis de investigación e hipótesis nula de las poblaciones. Las dos poblaciones son:

Población 1: parejas que participan en las tareas excitantes.
Población 2: parejas que participan en las tareas de control.

La hipótesis de investigación establecía que las parejas de la población 1 tendrían puntuaciones más altas que las parejas de la población 2 en cuanto a lo estimulante de las tareas (preguntas de control de manipulación): $\mu_1 > \mu_2$. Es decir, se trata de una hipótesis direccional. La hipótesis nula establecía que las parejas de la población 1 no tendrían puntuaciones más altas que las parejas de la población 2: $\mu_1 \leq \mu_2$.

2. **Determinar las características de la distribución comparativa.** Como ya observamos, la media de la distribución de diferencias de medias casi siempre es 0, ya que lo que nos interesa es saber si existe una diferencia mayor a 0 entre las dos poblaciones. La varianza poblacional, estimada a partir de las dos muestras de parejas, resulta ser 0,33 y 2,77.[1] La estimación combinada de la varianza poblacional es el promedio ponderado de las dos varianzas anteriores: 14/26 por 0,33 y 12/26 por 2,77. El resultado es 1,45. La varianza de cada distribución de medias, es decir, la estimación combinada dividida por el tamaño de cada muestra (1,45/15 y 1,45/13), es igual a

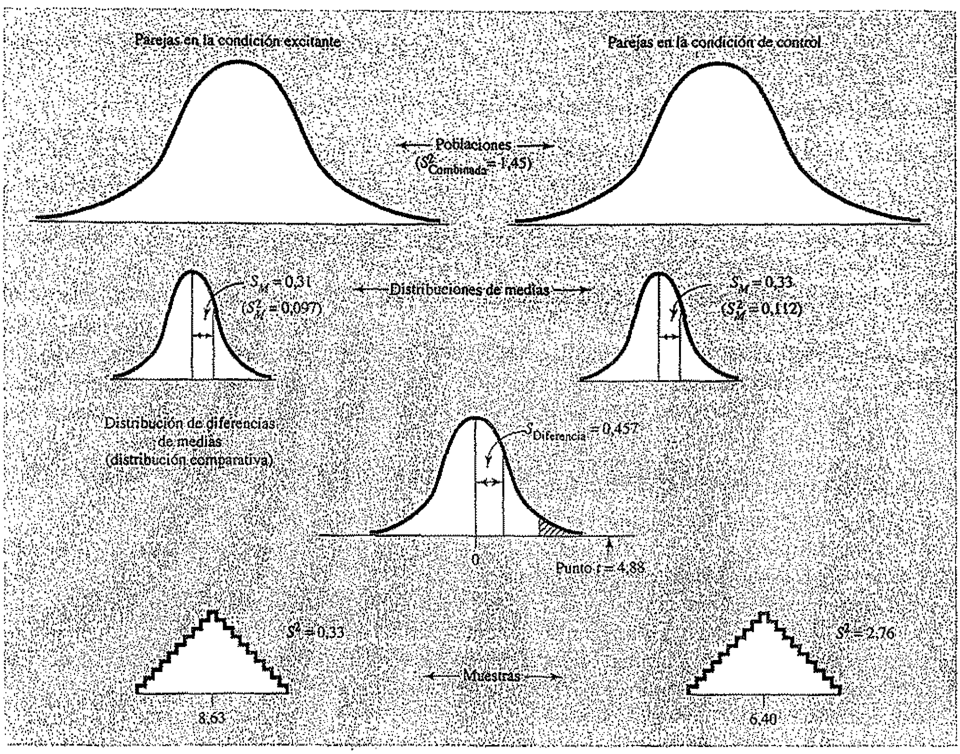


Figura 10-2. Distribuciones relacionadas con el ejemplo de control de manipulación en una prueba t para medias independientes. (**Fuente**: Norman & Aron, 1997).

[1] En este ejemplo, las varianzas estimadas de las dos poblaciones son sustancialmente diferentes. Esto genera objeciones en cuanto al supuesto de que ambas poblaciones tienen la misma varianza. Al final del capítulo, veremos el tema del supuesto de iguales varianzas poblacionales en forma general. No obstante, en este ejemplo, utilizar métodos alternativos que no requieran del supuesto produce resultados similares.

Tabla 10-1.
Prueba t para medias independientes correspondiente al control de manipulación de la excitación experimentada, comparando las condiciones de excitación con las de control.

Parejas en la condición de excitación			Parejas en la condición de control		
Registro	*Desvío de la media*	*Desvío cuadrático de la media*	*Registro*	*Desvío de la media*	*Desvío cuadrático de la media*
8,75	0,12	0,01	9,50	3,10	9,61
8,92	0,29	0,08	5,00	−1,40	1,96
9,50	0,87	0,76	4,83	−1,57	2,46
8,50	−0,13	0,02	8,42	2,02	4,08
8,17	−0,46	0,21	9,00	2,60	6,76
8,67	0,04	0,00	5,25	−1,15	1,32
8,17	−0,46	0,21	6,75	0,35	0,12
8,83	0,20	0,04	5,67	−0,73	0,53
9,17	0,54	0,29	6,17	−0,23	0,05
9,08	0,45	0,20	4,00	−2,40	5,76
8,75	0,12	0,01	6,50	0,10	0,01
7,08	−1,55	2,40	6,50	0,10	0,01
8,42	−0,21	0,04	5,67	−0,73	0,53
9,17	0,54	0,29			
8,33	−0,30	0,09			
Σ: 129,51		4,65	83,26		33,20

$M_1 = 8{,}63;\ S_1^2 = 4{,}65/14 = 0{,}33;\ M_2 = 6{,}40;\ S_2^2 = 33{,}20/12 = 2{,}77$

$N_1 = 15;\ gl_1 = N_1 - 1 = 14;\ N_2 = 13;\ gl_2 = N_2 - 1 = 12$

$gl_{\text{Total}} = gl_1 + gl_2 = 14 + 12 = 26$

$$S^2_{\text{Combinada}} = \frac{gl_1}{gl_{\text{Total}}}(S_1^2) + \frac{gl_2}{gl_{\text{Total}}}(S_2^2) = \frac{14}{26}(0{,}33) + \frac{12}{26}(2{,}77) = 0{,}54(0{,}33) + 0{,}46(2{,}77) = 0{,}18 + 1{,}27 = 1{,}45$$

$S^2_{M_1} = S^2_{\text{Combinada}}/N_1 = 1{,}45/15 = 0{,}097$

$S^2_{M_2} = S^2_{\text{Combinada}}/N_2 = 1{,}45/13 = 0{,}112$

$S^2_{\text{Diferencia}} = S^2_{M_1} + S^2_{M_2} = 0{,}097 + 0{,}112 = 0{,}209$

$S_{\text{Diferencia}} = \sqrt{S^2_{\text{Diferencia}}} = \sqrt{0{,}209} = 0{,}457$

t necesario para nivel 1%, gl=26, con prueba de una cola = 2,479

$t = (M_1 - M_2)/S_{\text{Diferencia}} = (8{,}63 - 6{,}40)/0{,}457 = 2{,}23/0{,}457 = 4{,}88$

Conclusión: se rechaza la hipótesis nula; se sostiene la hipótesis de investigación.

Fuente: Norman & Aron (1997).

0,097 y 0,112. Sumando estas varianzas obtenemos la varianza de la distribución de diferencias entre medias, 0,209. La raíz cuadrada de esta varianza, es decir, el desvío estándar de la distribución de diferencias de medias, es 0,457. La forma de la distribución comparativa será una distribución t con un total de 26 grados de libertad.

3. **Determinar el punto muestral de corte en la distribución comparativa, a partir del cual debería rechazarse la hipótesis nula.** El ejemplo requiere una prueba de una cola porque se predijo una dirección particular de la diferencia entre las dos poblaciones. Dado que todo el experimento depende del éxito de la manipulación, los investigadores son particularmente conser-

vadores al establecer el nivel de significación. Para una prueba de una cola con un nivel de 0,01, con 26 grados de libertad, la tabla *t* del apéndice B (tabla B-2) indica que necesitamos un *t* de al *menos* 2,479 para rechazar la hipótesis nula.

4. **Determinar el valor muestral en la distribución comparativa.** El punto *t* es la diferencia entre las dos medias muestrales de 2,23 (8,63 – 6,40 = 2,23) dividida por 0,457, el desvío estándar de la distribución de diferencias de medias. El resultado es una puntuación *t* de 4,88 (es decir, $t = 2{,}23/0{,}457 = 4{,}88$).

5. **Comparar los registros de los pasos 3 y 4 para decidir si se rechaza o no la hipótesis nula.** El valor *t* de 4,88, *correspondiente a la diferencia entre las medias de las dos condiciones*, es mayor que el *t* necesario de 2,479. Por lo tanto, los investigadores podrían rechazar la hipótesis nula con confianza. La hipótesis de investigación se sostiene: las parejas que participan de las condiciones de excitación informan que la actividad les resulta más excitante de lo que informan las parejas que participan en las condiciones de control. Por lo tanto, los investigadores podían confiar en que su manipulación experimental estaba funcionando en la forma deseada. (Por su puesto, el siguiente paso era ver si la manipulación experimental producía el aumento predicho de satisfacción marital y amor romántico).

Un dato interesante es que el análisis de los resultados completos de este estudio indicaron que las parejas en condiciones excitantes mostraron un aumento significativamente mayor de la satisfacción marital y del amor romántico entre antes y después de la actividad. (Es fácil interpretar estos resultados sabiendo que la manipulación experimental *sí* produce la sensación esperada).

Otros resultados de los estudios de Norman & Aron descartaron algunas explicaciones alternativas del efecto encontrado, y un estudio anterior (Reissman et al., 1993) descubrió el mismo efecto fuera del laboratorio en condiciones más realistas. Tomados en conjunto, estos estudios brindan una evidencia preliminar de que realizar actividades excitantes juntos puede ser una forma de aumentar la calidad marital de las parejas casadas.

Un segundo ejemplo de prueba *t* para medias independientes

Valenzuela (1997) comparó el cuidado maternal recibido por chicos pobres que estaban o no desnutridos. Una de sus medidas fueron índices acerca de la forma en que la madre ayudaba a su hijo en una tarea estándar de armado de rompecabezas (la observación se realizó durante visitas a las madres en sus casas, como parte de la investigación).

Los resultados obtenidos indicaron que las madres de los 43 niños adecuadamente alimentados tenían una media, en cuanto a la calidad de ayuda, de 33,1, y una varianza poblacional estimada de 201,64. Las madres de los 42 niños crónicamente desnutridos tenían una media de 27,0 en esta medida, con una varianza poblacional estimada de 134,56.

La figura 10-3 representa gráficamente la prueba *t* que compara la calidad de ayuda en las dos condiciones; la tabla 10-2 indica los valores y cálculos correspondientes. A continuación realizamos los cinco pasos de la prueba de hipótesis.

1. **Replantear el problema en función de hipótesis de investigación e hipótesis nula de poblaciones.** Las dos poblaciones son las siguientes:

Población 1: madres de niños pobres adecuadamente alimentados.
Población 2: madres de niños pobres crónicamente desnutridos.

La hipótesis de investigación establecía que las madres de la población 1 tendrían diferentes índices que las madres de la población 2 en cuanto a la calidad de ayuda brindada a sus hijos. Valenzuela predijo que la población 1 tendría índices más altos que los de la población 2. Sin embargo, siguiendo la práctica convencional en cuanto a estudios de este tipo, se utilizó una prueba de significación no direccional. (La ventaja de este tipo de prueba es que brinda la posibilidad de encontrar resultados significativos en la dirección opuesta a la predicción). Por lo tanto, la hipótesis de investigación realmente probada fue que las madres de la población 1 tendrían índices diferentes a los de las madres de la población 2; en símbolos, $\mu_1 \neq \mu_2$. La hipótesis nula establecía que las madres de la población 1 tendrían índices iguales a los de las madres de la población 2; $\mu_1 = \mu_2$.

2. **Determinar las características de la distribución comparativa**. Como es habitual, la media de la distribución de diferencias de medias será 0. La estimación combinada de la varianza poblacional es el promedio ponderado de las estimaciones de varianza poblacional realizados sobre la base de cada una de las dos muestras: 42/83 por 201,64 y 41/83 por 134,56. El resultado es 168,77. La varianza de cada distribución de medias, es decir, la estimación combinada dividida por el tamaño de cada muestra (168,77/43 y 168,77/42), es 3,92 y 4,02. Sumando los dos resultados anteriores obtenemos la varianza de la distribución de diferencias de medias, 7,94. La raíz cuadrada de esa varianza, es decir, el desvío estándar de la distribución de diferencias de medias, es 2,82. La forma de la distribución comparativa es una distribución *t* con un total de 83 grados de libertad.

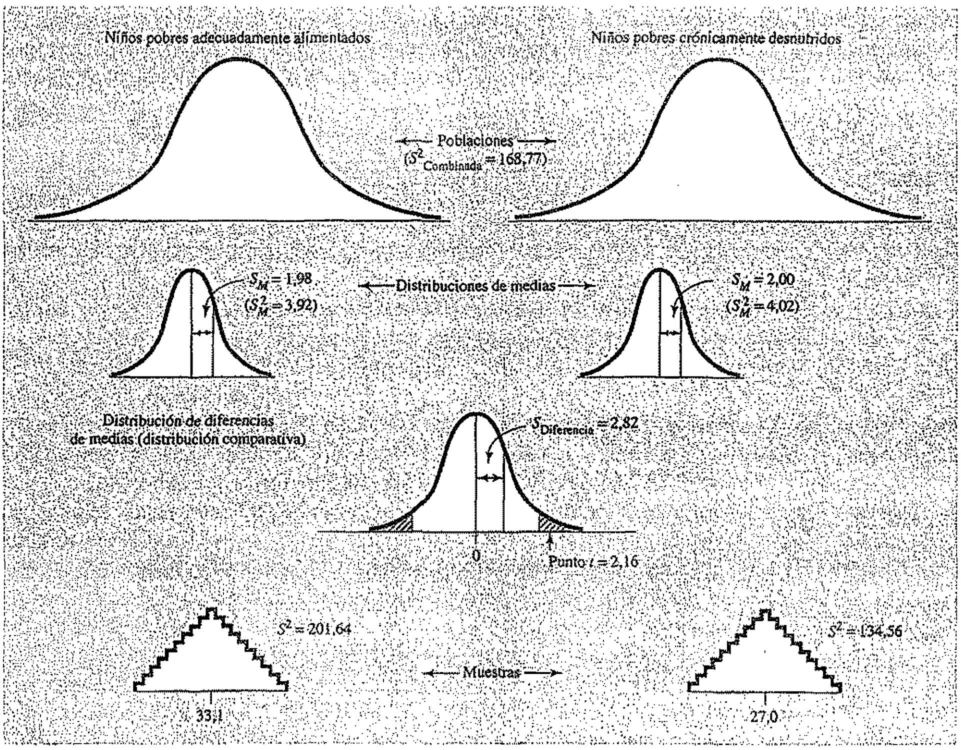


Figura 10-3. Distribuciones relacionadas con el ejemplo acerca de madres de niños pobres adecuadamente alimentados en comparación con madres de niños pobres crónicamente desnutridos. (**Fuente**: Valenzuela, 1997).

3. **Determinar el punto de corte en la distribución comparativa, a partir del cual debería rechazarse la hipótesis nula.** El punto de corte que necesitamos es el de una prueba de dos colas al nivel usual de 0,05, con 83 grados de libertad. Dado que la tabla t del apéndice B (tabla B-2) no incluye los datos para 83 grados de libertad, utilizamos el gl menor más cercano, que es 80. Obtenemos así un punto de corte de ±1,990 (En realidad son dos puntos simétricos: –1,99 y +1,99).

4. **Determinar el valor muestral en la distribución comparativa.** La puntuación t es la diferencia entre las dos medias muestrales dividida por el desvío estándar de la distribución de diferencias de medias. El resultado es un t de 2,16. (Es decir, $t = 6{,}1/2{,}82 = 2{,}16$).

5. **Comparar los valores de los pasos 3 y 4 para decidir si se rechaza o no la hipótesis nula.** La puntuación t de 2,16, correspondiente a la diferencia entre las medias de las dos condiciones, es más extremo que el punto t necesario de ±1,99. Por lo tanto, el investigador podría rechazar la hipótesis nula. Se sostiene la hipótesis de investigación: las madres de los niños adecuadamente alimentados brindan una mejor calidad de asistencia a sus niños que las madres de niños crónicamente desnutridos. (La tabla 10-7 indica los resultados completos de este estudio en la sección en la que tratamos el modo en que se describen los resultados de las pruebas t para medias independientes en las publicaciones científicas).

Un tercer ejemplo de prueba t para medias independientes

Aquí presentamos otro ejemplo, esta vez utilizando datos ficticios (de forma tal que podemos hacer que los números sean especialmente fáciles de manejar). Supongamos que un psicólogo especializado en rehabilitación ha desarrollado un nuevo programa de capacitación laboral para

Tabla 10-2.
Prueba t para medias independientes del estudio acerca de la calidad de asistencia brindada por madres de niños chilenos pobres adecuadamente alimentados, en comparación con madres de niños chilenos pobres crónicamente desnutridos.

Niños adecuadamente alimentados:

$$N_1 = 43;\ gl_1 = N_1 - 1 = 42;\ M_1 = 33{,}1;\ S_1^2 = 201{,}64$$

Niños crónicamente desnutridos:

$$N_2 = 42;\ gl_2 = N_2 - 1 = 41;\ M_2 = 27{,}0;\ S_2^2 = 134{,}56$$

$$gl_{Total} = gl_1 + gl_2 = 42 + 41 = 83$$

$$S^2_{Combinada} = \frac{gl_1}{gl_{Total}}(S_1^2) + \frac{gl_2}{gl_{Total}}(S_2^2) = \frac{42}{83}(201{,}64) + \frac{41}{83}(134{,}56)$$

$$= 0{,}51(201{,}64) + 0{,}49(134{,}56) = 1020{,}84 + 65{,}93 = 168{,}77$$

$$S^2_{M_1} = S^2_{Combinada}/N_1 = 168{,}77/43 = 3{,}92$$

$$S^2_{M_2} = S^2_{Combinada}/N_2 = 168{,}77/42 = 4{,}02$$

$$S^2_{Diferencia} = S^2_{M_1} + S^2_{M_2} = 3{,}92 + 4{,}02 = 7{,}94$$

$$S_{Diferencia} = \sqrt{S^2_{Diferencia}} = \sqrt{7{,}94} = 2{,}82$$

t necesario con nivel 5%, $gl = 83$ (utilizando un gl=80 de la tabla) y prueba de dos colas = 1,990

$$t = (M_1 - M_2)/S_{Diferencia} = (33{,}1 - 27{,}0)/2{,}82 = 6{,}1/2{,}82 = 2{,}16$$

Conclusión: se rechaza la hipótesis nula; se sostiene la hipótesis de investigación.

Fuente: Valenzuela (1997).

personas que no han sido capaces de mantener un empleo. Catorce personas acuerdan participar en el estudio, y el investigador escoge al azar siete de esos voluntarios para formar el grupo experimental que realizará el programa de capacitación especial. Los otros siete voluntarios formarán el grupo de control que realizará un programa de capacitación laboral ordinario. Después de finalizar los programas de capacitación (de ambos tipos), los 14 son ubicados en empleos similares.

Un mes después, se le pide al empleador de cada voluntario que califique el desempeño del participante utilizando una escala de 9 puntos. La tabla 10-3 indica los resultados ficticios y el análisis completo de la prueba *t*. La figura 10-4 representa gráficamente el análisis. Realicemos también el análisis paso a paso, siguiendo el procedimiento de prueba de hipótesis.

1. **Replantear el problema en función de hipótesis de investigación e hipótesis nula de poblaciones.** Las dos poblaciones son las siguientes:

Población 1: individuos que no podían mantener un empleo y que, po lo tanto, participan en el programa de capacitación laboral especial.

Población 2: individuos que no podían mantener un empleo y que, por lo tanto, participan en un programa de capacitación laboral ordinario.

Es posible que el programa especial tenga efectos positivos o negativos en comparación con el programa ordinario, por lo cual ambos resultados son de interés. Por lo tanto, la hipótesis de investigación establece que las medias de las dos poblaciones son diferentes: $\mu_1 \neq \mu_2$. Se trata de una hipótesis no direccional. La hipótesis nula establece que las medias de las dos poblaciones son iguales: $\mu_1 = \mu_2$.

2. **Determinar las características de la distribución comparativa.** La distribución de diferencias entre medias tendrá una media de 0, como es habitual. Determinamos su desvío estándar de la siguiente manera; a) calculando la varianza poblacional estimada sobre la base de cada muestra; b) calculando la estimación combinada; c) en el caso de cada población, dividiendo la estimación combinada por el tamaño de cada muestra para obtener la varianza de cada distribución de medias; d) sumando las varianzas de las dos distribuciones de medias para obtener la varianza de la distribución de diferencias de medias, y e) calculando la raíz cuadrada de esa varianza. Como lo indica la tabla 10-3, todo este proceso da como resultado un desvío estándar de 1,10. La forma de la distribución comparativa es una distribución *t* con un total de 12 grados de libertad.

3. **Determinar el punto de corte en la distribución comparativa, a partir del cual debería rechazarse la hipótesis nula.** Los investigadores utilizan el nivel de significación usual de 0,05 y una prueba de dos colas (ya que la hipótesis no es direccional). Buscando estos datos en la tabla *t*, en la línea correspondiente a 12 grados de libertad, descubrimos que necesitamos un punto *t* de al menos ±2,179.

4. **Determinar el valor muestral en la distribución comparativa.** La diferencia de medias dividida por el desvío estándar de la distribución de diferencias entre medias es una puntuación t de 2,73.

5. **Comparar los valores de los pasos 3 y 4 para decidir si se rechaza o no la hipótesis nula.** La puntuación *t* de 2,73 es más extrema que el 2,179 necesario. Por lo tanto, los investigadores rechazarían la hipótesis nula y concluirían que se sostiene la hipótesis de investigación: el nuevo programa de capacitación laboral es efectivo.

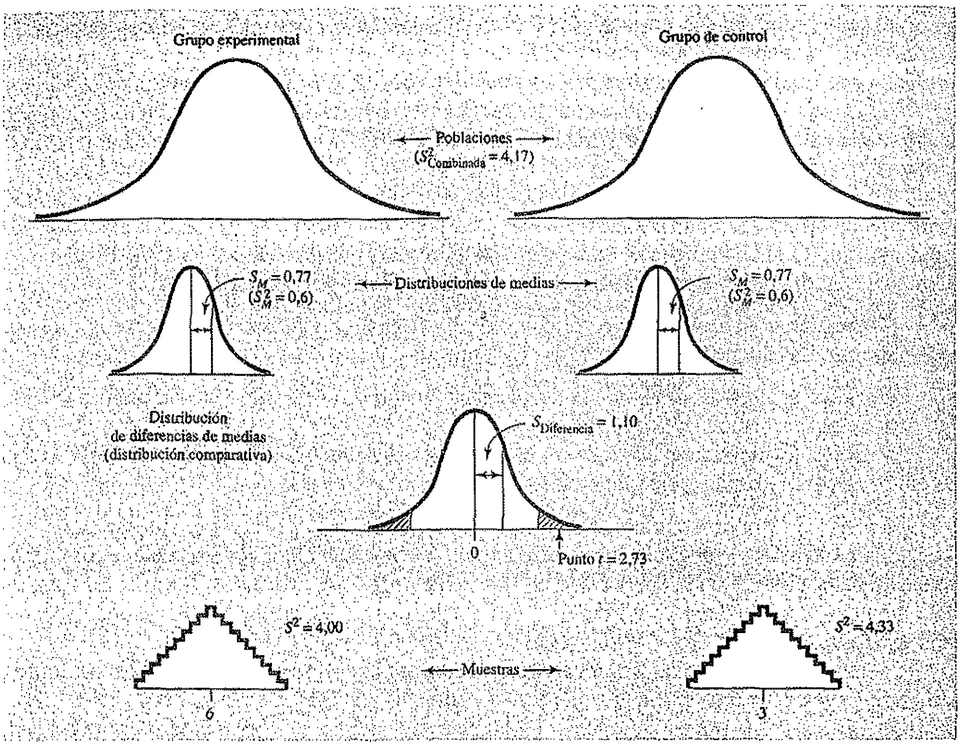


Figura 10-4. Distribuciones relacionadas con un experimento ficticio que prueba un nuevo programa de capacitación laboral.

Resumen de los pasos a seguir para la realización de una prueba *t* para medias independientes

La tabla 10-4 resume los pasos a seguir para realizar una prueba *t* para medias independientes. En el apéndice del capítulo se encuentran las fórmulas de cálculo, las cuales serán de gran utilidad si alguna vez es necesario calcular manualmente una prueba *t* de medias independientes (sin una computadora) para un estudio real con una gran cantidad de participantes. Sin embargo, para un mejor aprendizaje, recomendamos insistentemente que los ejercicios se resuelvan utilizando las fórmulas de definición y los procedimientos que se indican en la tabla 10-4.

SUPUESTOS DE LA PRUEBA *t* PARA MEDIAS INDEPENDIENTES

La primera suposición en una prueba *t* para medias independientes es igual a la suposición en cualquier prueba *t*: se supone que las distribuciones poblacionales son normales. En la práctica, esto implica un problema sólo si se considera que las dos poblaciones tienen distribuciones marcadamente asimétricas y en direcciones opuestas. En general, la prueba *t* se aplica bastante bien en la práctica aun cuando las formas de las distribuciones poblacionales sean moderadamente diferentes de la curva normal.

Tabla 10-3.
Cálculos de una prueba t para medias independientes correspondientes a un experimento de evaluación de la efectividad de un nuevo programa de capacitación laboral (utilizando la calificación de los empleadores) para personas que anteriormente no habían podido mantener sus empleos.

Grupo experimental (recibe el programa especial)			Grupo de control (Recibe el programa estándar)		
Registro	*Desvío de la media*	*Desvío cuadrático de la media*	*Registro*	*Desvío de la media*	*Desvío cuadrático de la media*
6	0	0	6	3	9
4	−2	4	1	−2	4
9	3	9	5	2	4
7	1	1	3	0	0
7	1	1	1	−2	4
3	−3	9	1	−2	4
6	0	0	4	1	1
Σ: 42	0	24	21	0	26

$M_1 = 6;\ S_1^2 = 24/6 = 4;\ M_2 = 3;\ S_2^2 = 26/6 = 4{,}33$

$N_1 = 7;\ gl_1 = N_1 - 1 = 6;\ N_2 = 7;\ gl_2 = N_2 - 1 = 6$

$gl_{\text{Total}} = gl_1 + gl_2 = 6 + 6 = 12$

$$S^2_{\text{Combinada}} = \frac{gl_1}{gl_{\text{Total}}}(S_1^2) + \frac{gl_2}{gl_{\text{Total}}}(S_2^2) = \frac{6}{12}(4) + \frac{6}{12}(4{,}33) = 0{,}5(4) + 0{,}5(4{,}33) = 2{,}00 + 2{,}17 = 4{,}17$$

$S^2_{M_1} = S^2_{\text{Combinada}}/N_1 = 4{,}17/7 = 0{,}60$

$S^2_{M_2} = S^2_{\text{Combinada}}/N_2 = 4{,}17/7 = 0{,}60$

$S^2_{\text{Diferencia}} = S^2_{M_1} + S^2_{M_2} = 0{,}60 + 0{,}60 = 1{,}20$

$S_{\text{Diferencia}} = \sqrt{S^2_{\text{Diferencia}}} = \sqrt{1{,}20} = 1{,}10$

t necesario para un nivel 5%, $gl = 12$, 5% y prueba de dos colas $= \pm 2{,}179$

$t = (M_1 - M_2)/S_{\text{Diferencia}} = (6{,}00 - 3{,}00)/1{,}10 = 3{,}00/1{,}10 = 2{,}73$

Conclusión: se rechaza la hipótesis nula; se sostiene la hipótesis de investigación.

En una prueba t para medias independientes existe una segundo supuesto muy importante, que ya hemos mencionado: se supone que las dos poblaciones tienen la misma varianza. (Aprovechamos este supuesto cuando promediamos las estimaciones de cada una de las muestras). Sin embargo, una vez más sucede que en la práctica la prueba t da resultados bastante precisos aun cuando existen diferencias considerablemente grandes entre las varianzas poblacionales, particularmente cuando existe la misma cantidad –o prácticamente la misma cantidad– de observaciones en las dos muestras. (¿Cómo sabemos que la prueba t se aplica adecuadamente a pesar de incumplimientos moderados de las presunciones? *Véase* en el cuadro 10-1 la descripción de lo que se denomina "Método de Montecarlo").

Sin embargo, la prueba t puede dar resultados bastante engañosos si a) los valores muestrales sugieren que las poblaciones son muy diferentes de lo normal, b) las varianzas son muy diferentes o c) coexisten ambos problemas. En esos casos, existen alternativas al procedimiento ordinario de prueba t, algunas de las cuales trataremos en el capítulo 15.

Tabla 10-4.
Pasos a seguir para la realización de una prueba t para medias independientes.

1. Replantear el problema en función de hipótesis de investigación e hipótesis nula de las poblaciones.
2. Determinar las características de la distribución comparativa.
   a) La media será 0.
   b) Calcular el desvío estándar.
      i) Calcular las varianzas poblacionales estimadas sobre la base de cada muestra (es decir, calcular dos estimaciones).
      ii) Calcular una estimación combinada de la varianza poblacional.
      $$S^2_{\text{Combinada}} = \frac{gl_1}{gl_{\text{Total}}}(S_1^2) + \frac{gl_2}{gl_{\text{Total}}}(S_2^2)$$
      $(gl_1 = N_1 - 1$ and $gl_2 = N_2 - 1$; $gl_{\text{Total}} = gl_1 + gl_2)$
      iii) Calcular la varianza de cada distribución de medias: $S^2_{M_1} = S^2_{\text{Combinada}}/N_1$ and $S^2_{M_2} = S^2_{\text{Combinada}}/N_2$
      iv) Calcular la varianza de la distribución de diferencias de medias:
      $S^2_{\text{Diferencia}} = S^2_{M_1} + S^2_{M_2}$
      v) Calcular el desvío estándar de la distribución de diferencias de medias:
      $S_{\text{Diferencia}} = \sqrt{S^2_{\text{Diferencia}}}$
   c) Determinar la forma: será una distribución t con gl_{Total} grados de libertad.
3) Determinar el punto de corte en la distribución comparativa, a partir del cual debería rechazarse la hipótesis nula
   a) Determinar los grados de libertad (gl_{Total}), el nivel de significación deseado, y las colas de la prueba (una o dos).
   b) Buscar el punto de corte apropiado en la tabla t. Si no aparece el gl exacto, se utiliza el gl inmediatamente inferior al buscado.
4) Determinar el valor muestral en la distribución comparativa: $t = (M_1 - M_2)/S_{\text{Diferencia}}$
5) Comparar los valores de los pasos 3 y 4 para decidir si se rechaza o no la hipótesis nula.

TAMAÑO DEL EFECTO Y POTENCIA DE LA PRUEBA *t* PARA MEDIAS INDEPENDIENTES

Tamaño del efecto

El tamaño del efecto en la prueba t para medias independientes es la diferencia entre las medias poblacionales dividida por el desvío estándar de la población de observaciones individuales. Se expresa bajo la fórmula,

$$d = \frac{\mu_1 - \mu_2}{\sigma} \tag{10-8}$$

Las reglas de Cohen (1988) de la prueba t para medias independientes son las mismas que en todas las situaciones que hemos tratado hasta ahora: 0,20 para un tamaño del efecto pequeño, 0,50 para un tamaño del efecto mediano y 0,80 para una gran tamaño del efecto.

Analicemos un ejemplo de cálculo de este tipo de tamaño del efecto. Supongamos que un psicólogo especializado en temas ambientales está trabajando en una ciudad con altos niveles de contaminación en el aire. El psicólogo planifica un estudio acerca de la cantidad de ejercicios resueltos en una prueba de creatividad durante un periodo de una hora. El estudio compara el desempeño en dos condiciones: en la condición experimental, cada participante realiza la prueba en una habitación con un purificador de aire especial; en la situación de control, cada participante realiza la prueba en una habitación sin el purificador de aire. El investigador espera que el grupo de control obtenga probablemente valores similares a otros que han realizado esta prueba en el pasado, es decir, con una media de 21, pero que el grupo experimental se desempeñe mejor y que tenga una media aproximadamente de 29. Se sabe por investigaciones anteriores que la prueba en cuestión tiene un desvío estándar de aproximadamente 10. Por lo tanto, $\mu_1 = 29$, $\mu_2 = 21$, y $\sigma = 10$. Dadas estas cifras, $d = (\mu_1 - \mu_2)/\sigma = (29 - 21)/10 = 0{,}80$, es decir, una gran tamaño del efecto.

Cuando se utiliza información de un estudio ya realizado, el tamaño del efecto se estima como la diferencia entre las medias muestrales dividida por la estimación combinada del desvío estándar poblacional (la raíz cuadrada de la estimación combinada de la varianza poblacional). Se expresa bajo la fórmula,

$$d = \frac{M_1 - M_2}{S_{\text{Combinada}}} \tag{10-9}$$

Analicemos el estudio de Valenzuela (1997) acerca de la calidad de la asistencia brindada por las madres a sus hijos. La media muestral de madres de niños adecuadamente alimentados era 33,1; la media muestral de madres de niños crónicamente desnutridos era 27,0. Calculamos que la estimación combinada de la varianza poblacional era 168,77; el desvío estándar era, por lo tanto, 12,99. La diferencia de medias era 6,1 y, al dividirla por 12,99, obteníamos un tamaño del efecto igual a 0,47, es decir, un tamaño del efecto mediano. La fórmula es la siguiente,

$$d = \frac{M_1 - M_2}{S_{\text{Combinada}}} = \frac{33{,}1 - 27{,}0}{12{,}99} = \frac{6{,}1}{12{,}99} = 0{,}47$$

Potencia

La tabla 10-5 indica la potencia aproximada correspondiente a un nivel de significación de 0,05, para tamaños del efecto pequeños, medianos y grandes, y para pruebas de una y dos colas. Analicemos nuevamente el ejemplo acerca de la psicología ambiental, en el que los investigadores esperaban un gran tamaño del efecto ($d = 0{,}80$). Supongamos que el investigador planifica realizar un estudio utilizando un nivel de 0,05, con una prueba de una cola y 10 participantes. Utilizando la tabla, el estudio tendría una potencia de 0,53, lo que implica que, aun si la hipótesis de investigación es realmente verdadera y tiene un gran tamaño del efecto, existe sólo un 53% de posibilidades de que el estudio resulte significativo.

Analicemos otro tipo de ejemplo. Supongamos que hemos leído un estudio que utiliza una prueba *t* para medias independientes, el cual tuvo un resultado no significativo utilizando un nivel de significación de 0,05 en una prueba de dos colas con 50 participantes en cada grupo. ¿Deberíamos concluir que, en realidad, no existe ninguna diferencia entre las poblaciones? La conclusión parece bastante injustificada, ya que la tabla 10-5 indica que el estudio tendría una potencia de sólo 0,17 para un tamaño del efecto pequeño. Lo anterior sugiere que si ese pequeño efecto de hecho sí existe en las poblaciones, el estudio no lo reflejaría. Por otro lado, también podemos concluir que si existe una verdadera diferencia entre las poblaciones, probablemente no es una gran diferencia, ya que la tabla 10-5 indica una potencia de 0,98 para un gran tamaño del efecto, lo que implica que si existiera un gran efecto, casi con seguridad sería reflejado por el estudio.

Cuadro 10-1. Los métodos de Montecarlo, o bien, cuando la matemática se convierte sólo en un experimento y la estadística depende de un juego de azar

El nombre Montecarlo, con el que se denominan ciertos métodos (por la famosa ciudad monegasca de veraneo y de juegos de azar), se adoptó hace sólo unos pocos años. Pero el método en sí mismo tiene su origen algunos siglos atrás, en la época en la que los matemáticos dejaban sus lápices o tizas y salían a intentar un experimento real para probar la interpretación particular de un problema de probabilidad. Por ejemplo, en 1777, Buffon describió en su *Essai d'Arithmétique morale*, un método de cálculo de la razón entre el diámetro de un círculo y su circunferencia, lanzando una aguja sobre una superficie plana con líneas paralelas. Presumiendo que la aguja cayera al azar en cualquier posición, uno podía calcular las posibilidades de que cayera en ciertas posiciones, como por ejemplo, la posibilidad de que tocara las líneas o no y de que cayera en ciertos ángulos. El término **Montecarlo** refleja, sin duda, la antigua interpretación de los matemáticos y estadísticos en cuanto a que muchos de sus problemas eran similares a aquellos que involucraban juegos de azar (recordemos a Pascal y al problema de los puntos descripto en el cuadro 5-2).

La utilización generalizada de los métodos de Montecarlo se hizo posible con el advenimiento de las computadoras, ya que la esencia de los estudios de Montecarlo es la interacción entre el azar y las probabilidades, lo que significa someter a prueba una gran cantidad de posibilidades. De hecho, la primera aplicación de los métodos de Montecarlo ocurrió en el campo de la física nuclear, dado que el comportamiento de las partículas, al ser esparcidas por un rayo de neutrones, es tan complicado y tan cercano a lo aleatorio que resolver el problema matemáticamente a partir de ecuaciones era prácticamente imposible. Sin embargo, simulando artificialmente las condiciones estadísticas de lo que esencialmente eran experimentos físicos, era posible comprender el mundo físico, o al menos aproximarse a él de manera más adecuada.

El alumno seguramente recordará el movimiento browniano que nos mostraban en las clase de química o física en la escuela secundaria. Su estudio es un buen ejemplo de un problema Montecarlo. Se trata, en líneas generales, de partículas atómicas, esta vez en un fluido, libres para hacer una cantidad casi ilimitada de cosas prácticamente al azar. De hecho, el movimiento browniano se ha comparado con la "caminata al azar" de un borracho: en cualquier momento, el borracho podría moverse en cualquier dirección. Pero el problema se simplifica limitando al borracho (o a la partícula) a una cuadrícula imaginaria.

Imaginemos la cuadrícula de las calles de una ciudad. Imaginemos también que hay una pared alrededor de la ciudad que el borracho no puede cruzar (de la misma manera que las partículas deben tener un límite y no pueden avanzar eternamente). En el límite (la pared), el borracho debe pagar una multa, que también varía al azar. El objetivo de este ejemplo es determinar cuánto cuesta el azar (todos los movimientos, así como también todas las consecuencias finales.) Por lo tanto, la cantidad de posibles recorridos es enorme.

El ejemplo de la caminata al azar nos lleva a la característica principal de los métodos Montecarlo: requieren la utilización de números aleatorios. Podemos encontrar una explicación acerca de estos números más adelante en el cuadro 15-1.

Volvamos ahora a lo que nos interesa, es decir, la utilización de los estudios Montecarlo para probar cuál será el resultado de los incumplimientos de ciertos supuestos en las pruebas estadísticas. Por ejemplo, la computadora puede crear dos poblaciones con medias idénticas, mientras que los otros parámetros son establecidos por el investigador estadístico de manera que violen algún supuesto importante. Las pobla- ciones podrían ser asimétricas hacia cierto lado, o bien, las dos poblaciones podrían tener varianzas diferentes.

Después se toman muestras aleatorias de cada una de estas dos extrañas poblaciones (recordemos, fueron inventadas por una computadora), se comparan las medias muestrales utilizando el procedimiento usual de prueba *t*, con las usuales tablas *t*, con todos los supuestos. Se selecciona una gran cantidad de tales pares de muestras, generalmente alrededor de 1.000, y se calcula una prueba *t* para cada par. La cuestión es: ¿Cuántas de esas 1.000 pruebas *t* serán significativas al nivel de significación del 5%? Lo ideal sería que el resultado sea aproximadamente del 5%, ó 50 de las 1.000. Pero ¿qué sucedería si el 10% (100) de esas pruebas, supuestamente a nivel 5%, resultara significativo? ¿Qué sucedería si fuera sólo el 1%? Si se dieran este tipo de resultados, entonces ese incumplimiento en particular de presunciones en la prueba *t* no podría ser tolerado. Pero de hecho, la mayoría de los incumplimientos (excepto los muy extremos), controlados con el método descripto, no crean grandes cambios en los valores *p*.

Los métodos Montecarlo son todo un suceso para la estadística, pero como todo, también tienen sus desventajas, y por lo tanto sus críticos. Uno de los inconvenientes es que el modo en que las poblaciones pueden violar las presunciones es casi ilimitado en cuanto a sus variaciones, y dado que incluso las computadoras tienen límites, los estudios Montecarlo son probados sólo en una serie representativa de esas variaciones. Otro inconveniente más específico es que existen buenas razones para pensar que algunas de las variaciones que no se analizan son mucho más semejantes a la vida real que aquellas que se han estudiado (*véase* en el capítulo 5 la controversia acerca de cuán común es realmente la curva normal). Finalmente, cuando intentamos decidir la utilización de un cálculo o prueba estadística en particular, en cualquier situación específica, no tenemos idea de la población de la cual proviene la muestra: ¿Es una población semejante a alguna de aquellas sobre las cuales se ha realizado un estudio Montecarlo o no? Saber simplemente que los estudios Montecarlo han demostrado que algunos cálculos y pruebas estadísticas son robustos a pesar de incumplimientos a distintos tipos de supuestos, no prueba que lo sean en cualquier situación determinada. Sólo nos da cierta esperanza en cuanto a que existan más posibilidades de que utilizar ese cálculo o prueba estadística sea seguro y justificable.

En todo caso, los estudios Montecarlo son un ejemplo perfecto del modo en que las computadoras han cambiado la ciencia. Shreider (1966) lo expresó de la siguiente manera:

> Las computadoras han producido una revolución única en la matemática. Mientras que anteriormente una investigación de un proceso aleatorio se consideraba completa tan pronto como fuera reducida a una descripción analítica, actualmente, en muchos casos es conveniente resolver un problema analítico reduciéndolo al proceso aleatorio correspondiente y luego simulando ese proceso (p. vii).

En otras palabras, en lugar de que la matemática nos ayude a analizar experimentos, son los experimentos los que nos están ayudando a analizar la matemática.

Tabla 10-5.
Potencia aproximada de estudios en los que se utiliza la prueba t para medias independientes, probando la hipótesis a un nivel de significación de 0,05.

Cantidad de participantes en cada grupo	Tamaño del efecto		
	Pequeño (0,20)	Mediano (0,50)	Grande (0,80)
Prueba de una cola			
10	0,11	0,29	0,53
20	0,15	0,46	0,80
30	0,19	0,61	0,92
40	0,22	0,72	0,97
50	0,26	0,80	0,99
100	0,41	0,97	*
Prueba de dos colas			
10	0,07	0,18	0,39
20	0,09	0,33	0,69
30	0,12	0,47	0,86
40	0,14	0,60	0,94
50	0,17	0,70	0,98
100	0,29	0,94	*

*Casi 1.
Nota: basado en Cohen (1988), pp. 28–39.

La potencia cuando los tamaños de las muestras no son iguales

La potencia es mayor cuando los participantes de un estudio se dividen en dos grupos iguales. Por ejemplo, un experimento con 10 personas en el grupo de control y 30 en el grupo experimental es mucho menos potente que uno con 20 personas en cada grupo.

Existe un problema práctico al calcular la potencia a partir de las tablas cuando los tamaños de muestra no son los mismos. Como en la mayoría de las tablas de potencia, la tabla 10-5 supone cantidades iguales en cada uno de los dos grupos. ¿Qué debemos hacer cuando las dos muestras tienen distinta cantidad de participantes? En lo que a la potencia respecta, ocurre que la **media armónica** de los dos tamaños desiguales de muestra nos indica el tamaño de muestra equivalente que tendríamos con dos muestras iguales. El tamaño de muestra que proviene de la media armónica se representa bajo la siguiente fórmula:

$$\text{Media armónica} = \frac{(2)(N_1)(N_2)}{N_1 + N_2} \qquad (10\text{-}10)$$

Analicemos un ejemplo extremo en el que hay 6 personas en un grupo y 34 personas en el otro. La media armónica es aproximadamente 10:

$$\text{Media armónica} = \frac{(2)(N_1)(N_2)}{N_1 + N_2} = \frac{(2)(6)(34)}{6+34} = \frac{408}{40} = 10{,}2$$

Por lo tanto, aunque tenemos un total de 40 participantes, el estudio tiene la potencia de un estudio con muestras iguales, de un tamaño de sólo 10 personas en cada grupo. (Es decir, un estudio con un total de 20 participantes habría tenido exactamente la misma potencia). Supongamos que el investigador está utilizando el nivel 0,05, una prueba de dos colas, y espera un gran tamaño del

Tabla 10-6.
Cantidad aproximada de participantes necesarios en cada grupo (suponiendo que las muestras son de igual tamaño) para obtener una potencia del 80% en una prueba *t* para medias independientes, probando la hipótesis a un nivel de significación de 0,05.

	Tamaño del efecto		
	Pequeño *(0,20)*	*Mediano* *(0,50)*	*Grande* *(0,80)*
Una cola	310	50	20
Dos colas	393	64	26

efecto. La tabla 10-5 indica que el estudio tendría una potencia de aproximadamente 0,39 (el número correspondiente a 10 participantes en cada grupo). Sin embargo, supongamos que el investigador hubiera podido organizar el estudio dividiendo los 40 participantes en dos grupos de 20. En ese caso, el estudio habría tenido una potencia de 0,69.

Planificación del tamaño de la muestra

La tabla 10-6 indica la cantidad aproximada de participantes necesarios para obtener una potencia del 80% para tamaños del efecto estimados pequeños, medianos o grandes, utilizando pruebas de una o dos colas, con un nivel de significación de 0,05 en todos los casos.[2] Supongamos que se planifica un estudio en el que se espera un tamaño del efecto mediano, y en el que se utilizará el nivel de significación de 0,05, con una prueba de una cola. Basándonos en la tabla 10-6, necesitaríamos 50 personas en cada grupo (100 en total) para tener una potencia del 80%. Si realizáramos un estudio con el mismo nivel de significación, pero en el que pudiéramos esperar un mayor tamaño del efecto, necesitaríamos sólo 20 personas en cada grupo (40 en total).

CONTROVERSIAS Y LIMITACIONES

Una vieja controversia se refiere a lo que usualmente llamamos el problema del "exceso de pruebas *t*". Las cuestiones básicas se presentan en todo tipo de prueba de hipótesis, no sólo en la prueba *t*. Sin embargo, analizamos el problema ahora porque tradicionalmente se ha tratado en este contexto.

Supongamos que se realizan una gran cantidad de pruebas *t* como parte del mismo estudio. Por ejemplo, podemos estar comparando dos grupos con cada una de 17 medidas diferentes, como pueden ser diferentes indicadores de memoria en una tarea en la que se emplea la capacidad de recordar varias sub-escalas de pruebas de inteligencia o diferentes aspectos de interacciones observados entre niños. Cuando se han realizado varias pruebas *t* en el mismo estudio, la posibilidad de que cualquiera de ellas resulte significativa a un nivel, digamos, del 5%, es realmente mayor al 5%. Si se realizan 100 comparaciones independientes, a un nivel del 5%, en promedio 5 de

[2] Cohen (1988, pp. 54-55) proporciona otras tablas que indican las cantidades necesarias de participantes para otros niveles de potencia además del 80%, para tamaños del efecto distintos de 0,20, 0,50 y 0,80, y para otros niveles de significación. Si es suficiente saber cuál es la cantidad aproximada, Dunlap y Myers (1997) han desarrollado una forma más corta de encontrar la cantidad aproximada de participantes necesarios para estudios que utilizan la prueba *t* para medias independientes. Para un 50% de potencia, la cantidad de participantes necesarios por grupo es aproximadamente $8/d^2 + 1$. Para un 80%-90% de potencia, $16/d^2 + 2$.

ellas serán significativas sólo por azar. Es decir, aproximadamente 5 serán significativas aun si no existiera ninguna diferencia real entre las poblaciones que las pruebas t están comparando.

La cuestión fundamental no es controvertida. Todo el mundo está de acuerdo con que existen inconvenientes en un estudio que incluye una gran cantidad de comparaciones. Todo el mundo está de acuerdo que en un estudio de ese tipo, si sólo unos pocos resultados son significativos, las diferencias reflejadas deberían ser revisadas muy cuidadosamente. La controversia surge en cuanto a cuán cuidadoso se debe ser y en cuanto a qué cantidad implica "sólo unos pocos". Una de las razones que da lugar a la controversia es que, en la mayoría de los casos, las muchas comparaciones que se realizan no son independientes, y la posibilidad de que una resulte significativa está relacionada con la posibilidad de que otra resulte significativa.

Veamos el siguiente ejemplo. Un estudio compara una muestra de abogados con una muestra de doctores con respecto a 100 rasgos de personalidad. Supongamos ahora que el investigador simplemente realiza 100 pruebas t. Si las 100 pruebas t fueran realmente independientes, esperaríamos que, en promedio, 5 resultaran significativas sólo por azar. De hecho, existen tablas para calcular con bastante precisión las chances de que cualquier cantidad determinada de pruebas t resulte significativa. De todos modos, el problema es que, en la práctica, estas 100 pruebas t no son independientes. Muchos de los distintos rasgos de personalidad probablemente estén correlacionados, como es el caso de las escalas que miden el dogmatismo y la confianza en sí mismos. Si los doctores y los abogados difieren en cuanto a dogmatismo, probablemente también tendrán diferencias en cuanto a confianza en sí mismos. Por lo tanto, ciertas series de comparaciones pueden tener más o menos probabilidades de resultar significativas por azar, de tal forma que 5 en 100 puede no ser un indicador preciso de cuántos resultados significativos esperar por azar.

Existe además otra complicación: en la mayoría de los casos, las diferencias en algunas de las variables son más importantes que en otras. Algunas comparaciones pueden probar directamente una teoría o la efectividad de algún procedimiento práctico, y otras pueden ser más "exploratorias". Esta complicación, junto con el problema de la falta de independencia, ha llevado a una variedad de soluciones conflictivas. En el capítulo 12 presentaremos algunas de esas soluciones cuando analicemos una situación relacionada con este tema, situación que surge en estudios que comparan más de dos grupos.

LA PRUEBA *t* PARA MEDIAS INDEPENDIENTES SEGÚN SE DESCRIBE EN PUBLICACIONES CIENTÍFICAS

Generalmente, una prueba t para medias independientes se describe en las investigaciones científicas por medio de las medias de las dos muestras (y a veces también de los desvíos estándar), además de la forma estándar de proporcionar los números t. Por ejemplo, "$t(38) = 4{,}72$, $p < 0{,}01$".

Los resultados del ejemplo de Norman y Aron (1997) podrían ser redactados de la siguiente manera: "En cuanto a los ítems del control de la manipulación, las puntuaciones fueron más altas en el caso de las parejas que formaban el grupo que realizó actividades excitantes ($M = 8{,}63$) que en el caso de las parejas del grupo de control ($M = 6{,}40$); $t(26) = 4{,}88$, $p < 0{,}01$, prueba de una cola."

Con frecuencia, los resultados de pruebas t para medias independientes se presentan en tablas. La tabla 10-7, indica los resultados de una cantidad de pruebas t para medias independientes pertenecientes al estudio de Valenzuela (1997) acerca de niños pobres desnutridos o no. La cuarta línea a partir de abajo indica los resultados de la parte del estudio que hemos utilizado como ejemplo en este capítulo.

Tabla 10-7.
Comparación de grupos adecuadamente alimentados y crónicamente desnutridos según medidas demográficas y de relación entre madre e hijo.

	Grupo adecuadamente alimentado ($n = 43$)		Grupo crónicamente desnutrido ($n = 42$)			
Medida	*M*	*SD*	*M*	*SD*	valor *t*	*p*
Materna y familiar						
Ingresos totales	45,30	9,0	44,7	10,0	0,30	0,77
Tamaño familiar	5,7	2,2	5,2	1,4	1,18	0,24
Cantidad de hermanos	2,6	0,8	2,8	0,8	0,85	0,39
Edad del padre	31,4	5,6	29,6	6,9	1,29	0,20
Educación del padre	7,2	2,8	6,8	2,9	0,64	0,52
Edad de la madre	28,6	4,7	27,6	5,7	0,84	0,40
Educación de la madre	7,0	2,6	6,1	2,9	1,39	0,17
Peso de la madre (kg.)	59,2	10,0	53,3	10,0	2,5	0,01
Estatura de la madre (cm.)	153,6	5,6	150,5	6,4	2,31	0,02
Sensibilidad materna	5,63	2,4	2,1	1,5	8,14	0,0001
Cuestionario sobre salud (registro total)	10,8	3,9	10,7	5,3	0,08	0,93
Escala de adaptación marital (registro total)	84,7	26,6	73,8	32,9	1,59	0,11
Niño						
Edad (meses)	18,5	1,4	18,4	1,5	0,33	0,74
Peso	103,72	7,3	81,6	3,9	17,40	0,0001
Estatura	98,3	2,8	92,9	3,3	7,93	0,0001
Funcionamiento madre-hijo						
Sensibilidad materna	7,7	3,8	7,3	3,6	0,58	0,59
Control materno	3,4	4,1	4,2	3,7	0,98	0,36
Insensibilidad materna	2,8	2,9	2,4	3,1	0,67	0,53
Cooperación del niño	7,9	4,5	6,8	4,1	0,12	0,24
Sumisión compulsiva del niño	1,2	3,2	2,1	3,8	0,12	0,26
Dificultad del niño	2,5	2,8	3,5	3,4	0,39	0,71
Pasividad del niño	2,2	2,7	2,4	2,6	0,36	0,72
Resolución de problemas						
Apoyo materno	37,9	10,6	30,54	7,9	3,62	0,001
Calidad de asistencia materna	33,1	14,2	27,0	11,6	2,16	0,03
Competencia social del niño	19,9	4,1	15,7	3,8	4,78	0,0001
Demostración de poder del niño	15,6	5,3	12,5	4,5	2,85	0,006
Sentimientos negativos del niño	7,4	4,2	9,4	4,8	2,00	0,050

[a]Las medidas de peso y estatura de los niños están indicadas en forma de porcentajes del peso y medidas según la edad, conforme a las normas del Centro Nacional de Estadísticas Sanitarias.

Fuente: Valenzuela, M. (1997), tab. 1. "Sensibilidad materna en una sociedad en desarrollo: el contexto de la pobreza urbana y la desnutrición infantil crónica". *Psicología de Desarrollo [Developmental Psychology]*, 33, 845–855. Copyright, 1997, por la Asociación Americana de Psicología [American Psychological Association]. Reimpreso con autorización.

La tabla 10-8 es otro ejemplo, tomado de un estudio realizado por Frisch, Shamsuddin y Kurtz (1995), en el que 293 mujeres estudiantes de medicina en Malasia fueron entrevistadas acerca de sus opiniones en cuanto a fumar, y en cuanto a si fumaban o no los miembros de sus familias y sus amigos. La tabla compara aquellas estudiantes que tienen hermanos fumadores con aquéllas cuyos hermanos no fuman. (La publicación no explica cómo se resolvía el problema de que la persona tuviera dos hermanos, uno fumador y otro no fumador). Las medidas fueron: conocimiento (de los riesgos para la salud ocasionados por estar rodeado de fumadores), actitud (hacia estar rodeado de fumadores), esfuerzos (para evitar estar rodeado de fumadores) y responsabilidad como médico (de informar a los pacientes sobre los riesgos para la salud ocasionados por estar rodeado de fumadores). En todas las escalas, el puntaje estaba establecido de tal forma que el más alto estuviera a favor de fumar. Los valores más bajos significaban mayor preocupación acerca de los riesgos para la salud.

La primera línea de la tabla indica que aquellas que tenían un hemano fumador presentaban valores más altos en la escala de conocimiento, lo que significa que esas estudiantes tenían menos conocimiento acerca de los riesgos de estar con fumadores. La segunda línea indica que aquellas que tenían un hermano fumador tenían una actitud más positiva hacia estar con fumadores (es decir, no consideraban que era una causa de riesgo para la salud).

Es importante notar que algunos de estos resultados no fueron significativos. ¿Cuál debería ser la conclusión? Analicemos lo que piensan las estudiantes acerca de la responsabilidad como médico. En esta comparación, había 41 estudiantes con hermanos fumadores y 73 con hermanos no fumadores. Aplicando la fórmula de la media armónica observamos que, en lo que respecta al cálculo de la potencia, hay 52,5 participantes por grupo. Es decir,

$$\text{Media armónica} = \frac{(2)(N_1)(N_2)}{N_1 + N_2} = \frac{(2)(41)(73)}{41+73} = \frac{5.986}{114} = 52{,}5$$

Una vez que sabemos qué tamaño de muestra utilizar, podemos buscar la potencia en la tabla 10-5 buscando la hilera de 50 participantes (el número más cercano a 52,5 en la tabla) para una prueba

Tabla 10-8.
Prueba *t* para medias, acerca del conocimiento, las actitudes y el esfuerzo relacionados con el hecho de ser fumador pasivo, según la condición de fumador y con respecto al grupo total y a hombres y mujeres separadamente.

	Hermano fumador	**Hermano no fumador**	**Valor *t***	**Sig.**
Grupo total	*N* =	*N* =		
Conocimiento	2,03 (96)	1,88 (140)	2,61	0,01
Actitud	1,95 (94)	1,70 (137)	3,29	0,001
Esfuerzos	2,36 (92)	2,23 (133)	1,88	0,061
Resp. Médico*	1,78 (95)	1,61 (142)	2,02	0,04
Hombres				
Conocimiento	2,15 (54)	1,92 (69)	2,97	0,004
Actitud	2,08 (54)	1,83 (67)	2,12	0,036
Esfuerzos	2,50 (52)	2,31 (66)	1,87	0,064
Resp. Médico*	1,81 (54)	1,65 (69)	1,27	0,207
Mujeres				
Conocimiento	1,87 (42)	1,85 (71)	0,30	0,767
Actitud	1,77 (40)	1,57 (70)	2,43	0,018
Esfuerzos	2,17 (40)	2,15 (67)	0,26	0,797
Resp. Médico*	1,76 (41)	1,58 (73)	1,51	0,136

*Responsabilidad como médicos

Fuente: Frisch, Shamsuddin & Kurtz (1995).

de dos colas. Descubrimos que la potencia del estudio, para resultar significativa con un pequeño tamaño del efecto, es sólo de 0,17. Por otro lado, la potencia del estudio en el caso de un tamaño del efecto mediano es 0,70, y en el de un tamaño del efecto grande es 0,98. Así, si en realidad tener un hermano fumador produce un pequeño efecto, dicho efecto probablemente no habría sido reflejado por el estudio. Por otro lado, supongamos que en realidad había un efecto mediano de ese tipo; en ese caso, el resultado del estudio probablemente habría sido significativo; y si el efecto fuera grande, casi con seguridad el estudio habría resultado significativo. Por lo tanto, con bastante confianza podemos inferir de este estudio que el hecho de tener un hermano fumador probablemente no produce una gran diferencia en las opiniones de las mujeres estudiantes de medicina de Malasia, en cuanto a la responsabilidad del médico de informar a sus pacientes acerca de los riesgos de estar con fumadores. Pero no podemos concluir que no podría haber un pequeño efecto en ese sentido.

RESUMEN

Una prueba *t* para medias independientes se utiliza para realizar pruebas de hipótesis con dos muestras de observaciones. La diferencia principal con una prueba *t* para una sola muestra, o una prueba *t* para medias dependientes, es que la distribución comparativa es una distribución de diferencias entre medias muestrales. Esta distribución puede considerarse construida en dos pasos: cada población de individuos produce una distribución de medias y luego se crea una nueva distribución de diferencias entre pares de medias tomadas de esas dos distribuciones de medias.

La distribución de diferencias de medias tiene una media de 0, y es una distribución *t* con el total de los grados de libertad de las dos muestras. El desvío estándar se calcula en varios pasos: a) se utiliza cada muestra para estimar la varianza poblacional; b) se supone que ambas poblaciones tienen la misma varianza, y se realiza una estimación combinada sacando un promedio ponderado de las dos estimaciones (multiplicando cada estimación por la proporción con que contribuye su muestra a los grados totales de libertad y sumando los resultados); c) se divide la estimación combinada por la cantidad de observaciones de cada muestra para obtener la varianza de la distribución de medias de cada población; d) se suman esas dos varianzas para obtener la varianza de la distribución de diferencias de medias, y e) se calcula la raíz cuadrada.

Los supuestos en la prueba *t* para medias independientes son las siguientes: las dos poblaciones están normalmente distribuidas y tienen la misma varianza. Sin embargo, la prueba *t* otorga resultados bastante precisos aun cuando la situación real sea moderadamente diferente de lo que indican los supuestos.

El tamaño del efecto de una prueba *t* para medias independientes es la diferencia entre las medias dividida por el desvío estándar. La potencia es mayor cuando los tamaños de las muestras de los dos grupos son iguales. Cuando no lo son, se utiliza la media armónica de los dos tamaños muestrales para buscar la potencia en las tablas.

Cuando se realizan demasiadas pruebas de significación en el mismo estudio, como en el caso de una serie de pruebas *t* que comparan dos grupos con respecto a varias medidas, la posibilidad de que cualquiera de las comparaciones resulte significativa por azar al nivel del 0,05 es mayor a 0,05. La forma de adaptación de los cálculos para resolver este problema es controvertida, aunque la mayoría está de acuerdo con que en una situación de ese tipo los resultados deberían ser interpretados con mucho cuidado.

Cuando las publicaciones científicas informan acerca de pruebas *t* para medias independientes, el investigador usualmente incluye los grados de libertad, el punto *t* y el nivel de significación. También pueden informarse los resultados de estas pruebas a través de una tabla.

Términos clave

- Distribución de diferencias de medias.
- Media armónica.
- Estimación combinada de la varianza poblacional ($S^2_{\text{Combinada}}$).
- Desvío estándar de la distribución de diferencias de medias ($S_{\text{Diferencia}}$).
- Prueba t para medias independientes.
- Varianza de la distribución de diferencias entre medias ($S^2_{\text{Diferencia}}$).
- Promedio ponderado.

Ejercicios

Los ejercicios implican la realización de cálculos (con la ayuda de una calculadora). La mayoría de los problemas estadísticos reales se resuelven por computadora, pero aunque exista la posibilidad de utilizarla, es conveniente realizar estos ejercicios manualmente para incorporar el método de trabajo.

Para adquirir práctica en la utilización de una computadora, para resolver problemas estadísticos, se puede utilizar la sección de computación de cada capítulo, publicada en la *Guía de estudio y libro de tareas de computación para el alumno [Student's Study Guide and Computer Workbook]* que acompaña este libro.

Todos los datos de esta sección son ficticios (a menos que se especifique lo contrario)

Las respuestas a los ejercicios de la serie I se encuentran al final del libro.

SERIE I

1. a) Explique en qué casos utilizaría una prueba t para medias dependientes y cuándo una prueba t para medias independientes.
b) Invente un ejemplo sobre cada tipo de estudio que no esté en el libro ni se haya dado en clase .

2. Para cada uno de los siguientes experimentos, a) decida si la diferencia entre las condiciones es estadísticamente significativa a nivel 0,05 (con prueba de dos colas), b) determine el tamaño del efecto y c) la potencia aproximada (según la tabla 10-5).

	Grupo experimental			Grupo de control		
	N	M	S^2	N	M	S^2
i)	30	12,0	2,4	30	11,1	2,8
ii)	20	12,0	2,4	40	11,1	2,8
iii)	30	12,0	2,2	30	11,1	3,0

3. Un psicólogo social que estudia la comunicación masiva divide al azar a 82 voluntarios en dos grupos experimentales. A 61 se los instruyó para que se informaran durante un mes sólo a través de la televisión, y a 21 para que se informaran durante un mes sólo a través de la radio. (¡Es un misterio por qué el investigador no asignó iguales cantidades de voluntarios a las dos condiciones!). Finalizado el mes, todos los participantes fueron puestos a prueba con respecto a su conocimiento acerca de varias cuestiones políticas. El investigador no tenía ninguna predicción sobre cuál era la fuente de información que haría que las personas estuvieran mejor informadas, es decir, el investigador sólo predijo que habría algún tipo de diferencia. Los resultados del estudio fueron los siguientes: el grupo que se informó por TV: $M = 24$, $S^2 = 4$; el grupo que se informó por radio: $M = 26$, $S^2 = 6$. Utilizando el nivel 0,01, ¿cuál debería ser la conclusión del psicólogo social? a) Realice los cinco pasos de la prueba de hipótesis. b) Calcule el tamaño del efecto. c) Ilustre su respuesta con un diagrama similar al de las figuras 10-2 a 10-4. d) Explique sus respuestas a alguien que nunca ha tomado un curso de estadística.

4. Un psicólogo especializado en educación estaba interesado en saber si la utilización del nombre del alumno en un cuento afectaba

la concentración del niño mientras leía. Se asignaron seis niños al azar para leer un cuento en condiciones normales (utilizando nombres como Dick y Jane). Otros cinco niños leyeron versiones del mismo cuento, pero con el nombre de cada uno de ellos en lugar del de uno de los personajes de la historia. El investigador midió cuidadosamente cuánto tardaba cada niño en leer el cuento. Los resultados aparecen más adelante. Utilizando el nivel 0,05, ¿diría que incluir el nombre del niño produce alguna diferencia? a) Realice los cinco pasos de la prueba de hipótesis. b) Calcule el tamaño del efecto y la potencia. c) Ilustre su respuesta con un diagrama similar al de las figuras 10-2 a 10-4. d) Explique su respuesta a alguien que comprende la prueba t para medias dependientes pero no conoce nada acerca de la prueba t para medias independientes.

Cuento normal		**Cuento con el nombre del niño**	
Alumno	*Tiempo de lectura*	*Alumno*	*Tiempo de lectura*
A	2	G	4
B	5	H	16
C	7	I	11
D	9	J	9
E	6	K	8
F	7		

5. Van Aken y Asendorpf (1997) realizaron un estudio a 139 niños alemanes de 12 años de edad. Todos los niños completaron un cuestionario general sobre autovaloración y fueron entrevistados con respecto al apoyo que recibían de sus madres, padres y compañeros de clase. Luego, los investigadores compararon el nivel de autovaloración entre aquellos con altos y bajos niveles de apoyo para cada tipo de apoyo recibido. Los investigadores informaron lo siguiente:

> Se descubrió una menor autovaloración general en los niños con madres que les brindaban un bajo nivel de apoyo ($t(137) = 4{,}52$, $p < 0{,}001$, $d = 0{,}78$) y con padres que les brindaban un bajo nivel de apoyo ($t(137) = 4{,}03$, $p < 0{,}001$, $d = 0{,}69$) [...] También se descubrió una autovaloración menor en niños que sólo tenían compañeros que le brindaban un bajo nivel de apoyo ($t(137) = 2{,}04$, $p < 0{,}05$, $d = 0{,}35$).

Explique el significado de estos resultados a una persona que nunca ha tomado un curso de estadística. (Asegúrese de incluir el tema del tamaño del efecto y la potencia. Al calcular la potencia, puede suponer que los dos grupos de cada comparación tenían aproximadamente el mismo tamaño de muestra).

6. ¿Cuál es la cantidad aproximada de participantes necesarios para tener una potencia del 80% en cada uno de los siguientes estudios planificados, suponiendo que existe la misma cantidad de participantes en los dos grupos y utilizando en todos el nivel de significación de 0,05? (Asegúrese de calcular la cantidad total de participantes necesarios, no sólo la cantidad necesaria para cada grupo).

	Esperado		**Esperado**	
Estudio	μ_1	μ_2	σ	*Colas*
a	107	149	84	1
b	22,5	16,2	31,5	2
c	14	12	2,5	1
d	480	520	50	2

SERIE II

1. Explique con sus propias palabras cómo determinaría la varianza de la distribución de diferencias de medias (ilustre su respuesta con un diagrama que represente gráficamente todas las distribuciones involucradas).

2. Para cada uno de los siguientes experimentos, a) decida si la diferencia entre las distintas condiciones es estadísticamente significativa al nivel 0,05 (prueba de dos colas) y b) determine el tamaño del efecto (d) y la potencia aproximada (a partir de los datos de la tabla 10-5).

	Grupo experimental			**Grupo de control**		
	N	M	S^2	N	M	S^2
i)	10	604	60	10	607	50
ii)	40	604	60	40	607	50
iii)	10	604	20	40	607	16

3. Un psicólogo desarrolla la teoría de que las personas pueden escuchar mejor después de comer una comida abundante. Se dividieron al

azar 6 individuos para comer una comida abundante o una comida frugal. Después de comer, se probó el sentido auditivo de los participantes. Los valores observados de capacidad auditiva (los números más altos indican mayor capacidad) aparecen más adelante. Utilizando el nivel 0,05, ¿diría que los resultados sostienen la teoría del psicólogo? a) Realice los cinco pasos de la prueba de hipótesis. b) Calcule el tamaño del efecto y la potencia. c) Ilustre su respuesta con un diagrama semejante al de las figuras 10-2 a 10-4. d) Explique sus respuestas a alguien que nunca ha asistido a un curso de estadística.

Grupo comida abundante		**Grupo comida frugal**	
Sujeto	*Capacidad auditiva*	*Sujeto*	*Capacidad auditiva*
A	22	D	19
B	25	E	23
C	25	F	21

4. Veinte estudiantes asignados al azar a un grupo experimental reciben un programa de instrucción; 30 participantes de un grupo de control no lo reciben. Después de 6 meses, se prueba a ambos grupos en cuanto a sus conocimientos. El grupo experimental tiene una media de 38 en la prueba (con un desvío estándar poblacional estimado de 3); el grupo de control tiene una media de 35 (con un desvío estándar poblacional estimado en 5). Utilizando el nivel 0,05, ¿cuál debería ser la conclusión del experimentador? a) Realice los cinco pasos de la prueba de hipótesis. b) Calcule el tamaño del efecto y la potencia. c) Ilustre su respuesta con un diagrama similar al de las figuras 10-2 a 10-4. d) Explique sus respuestas a alguien que comprende la prueba *t* para medias dependientes pero no conoce nada sobre la prueba *t* para medias independientes.

5. En un estudio acerca del efecto de los colores para calmar la angustia, se compararon las puntuaciones en una prueba de angustia de participantes que completaron la prueba impresa en papel amarillo tenue, con las puntuaciones de participantes que completaron la prueba en papel verde chillón. Las puntuaciones de los cinco participantes que completaron la prueba impresa en papel amarillo fueron 17, 19, 28, 21 y 18. Las puntuaciones de los cuatro participantes que completaron la prueba en papel verde fueron 20, 26, 17 y 24. Utilizando el nivel 0,05, una cola (prediciendo menores puntuaciones de angustia con el papel amarillo), ¿cuál debería ser la conclusión del investigador? a) Realice los cinco pasos de la prueba de hipótesis. b) Calcule el tamaño del efecto y la potencia. c) Ilustre su respuesta con un diagrama semejante al de las figuras 10-2 a 10-4. d) Explique sus respuestas a alguien que comprende la prueba *t* para medias dependientes pero que no conoce nada sobre la prueba *t* para medias independientes.

6. Escudero, Rogers y Gutierrez (1997) filmaron en video a 30 parejas que discutían un problema marital en su laboratorio. Las cintas de video fueron luego evaluadas sistemáticamente según varios aspectos relacionados con la comunicación entre la pareja, tales como dominación y calidad positiva o negativa del sentimiento (emoción) expresado entre ellos. Uno de los principales intereses del estudio era comparar parejas que estaban teniendo problemas de relación con aquellas que no los tenían. Las 18 parejas del grupo que tenía problemas fueron reclutadas de entre aquellos que habían recurrido por ayuda a una clínica marital; se los llamaba el grupo "clínico". El grupo que no tenía problemas fue reclutado a través de anuncios, y se los denominaba el grupo "no clínico". (Los dos grupos, de hecho, tuvieron valores drásticamente diferentes en una prueba estándar de satisfacción marital). La tabla 10-9 presenta algunos de los resultados. (Se pueden ignorar las flechas y signos más y menos que están relacionados con la manera en que ellos clasificaron la interacción. Además, ignore la nota al pie sobre "transformación arco-seno"; explicaremos este concepto en el capítulo 15). Concentrándose en la dominación y en la sumisión, explique qué significan estos resultados a una persona que nunca ha asistido a un curso sobre estadística. (Asegúrese de tratar el tema del tamaño del efecto y la potencia. Al calcular la potencia, puede suponer que los dos

Tabla 10-9.
Diferencias de razones básicas para parejas clínicas y no clínicas acerca del control en la relación y el sentimiento no verbal expresado en proporciones (*SD* en paréntesis).

	Situación de pareja		Diferencias intergrupales
	Media clínica	*Media no-clínica*	*t*
Dominación (↑)	0,452 (107)	0,307 (0,152)	3,06*
Nivelación (→)	0,305 (0,061)	0,438 (0,065)	5,77**
Sumisión (↓)	0,183 (0,097)	0,226 (0,111)	1,12
Códigos dobles	0,050 (0,028)	0,024 (0,017)	2,92*
Sentimiento positivo (+)	0,127 (0,090)	0,280 (0,173)	3,22*
Sentimiento negativo (–)	0,509 (0,192)	0,127 (0,133)	5,38**
Sentimiento neutro (0)	0,344 (0,110*)	0,582 (0,089)	6,44**
Códigos dobles (+/–)	0,019 (0,028)	0,008 (0,017)	2,96*

Nota: las proporciones de cada código de control y de cada sentimiento fueron convertidas utilizando la transformación arco-seno para utilizarlas en comparaciones intergrupales $*p < 0{,}01$; $**p < 0{,}001$; ($gl = 28$).

Fuente: Escudero, V., Rogers, L. E. & Gutierrez, E. (1997), tab. 3. "Patrones de control en la relación y de sentimiento no verbal en parejas clínicas y no clínicas". *Revista Científica de Relaciones Sociales y Personales [Journal of Social and Personal Relationships]*, 14, 5–29. Copyright © 1997 por Sage Publications, Inc. Reimpreso con autorización de Sage Publications.

grupos de cada comparación tenían aproximadamente el mismo tamaño muestral).

7. ¿Quiénes tienen primeros nombres más largos, los hombres o las mujeres? Tome un directorio telefónico y utilice los números aleatorios que le proporcionamos a continuación para seleccionar una página. En la primera página (página 12), busque el primer nombre claramente femenino y anote la cantidad de letras de ese nombre. Haga lo mismo 16 veces (busque la página correspondiente al número indicado, etc.). Después busque la cantidad de letras de 16 nombres masculinos. (Tendrá que excluir aquellos nombres cuyo género no esté seguro). Calcule una prueba *t* para medias independientes utilizando esas dos muestras. (Asegúrese de destacar a qué ciudad pertenece el directorio telefónico que utilizó).

12, 79, 10, 97, 53, 74, 15, 55, 41, 128, 57, 93, 94, 31, 68, 516, 60, 56, 7, 93, 43, 91, 57, 58, 38, 120, 14, 38, 57, 743, 98, 471, 38, 66, 20, 32, 60, 43, 78, 29, 39, 17, 31, 12, 61, 100, 80, 35, 31, 99, 22

Apéndice del capítulo: fórmulas de cálculo optativas de la prueba *t* para medias independientes

El procedimiento descripto en este capítulo, sobre la base de fórmulas de definición, involucra cálculos de a) la media (M) de cada muestra, b) la varianza poblacional estimada (S^2) sobre la base de los valores de cada muestra, c) la estimación combinada de la varianza poblacional ($S^2_{\text{Combinada}}$), d) la varianza de la distribución de medias (S^2_M) de cada población, e) la varianza y el desvío estándar de la distribución de diferencias entre medias ($S^2_{\text{Diferencia}}$ y $S_{\text{Diferencia}}$) y f) el punto t.

Los primeros dos pasos (calcular M y S^2 para cada muestra) pueden calcularse utilizando las fórmulas de cálculo de los capítulos 2 y 9; estos cálculos estadísticos por lo general están disponibles directamente en las calculadoras. Existe, sin embargo, una fórmula especial de cálculo que combina los pasos c) a e) para obtener el desvío estándar de la distribución de diferencias entre medias ($S_{\text{Diferencia}}$):

$$S_{\text{Diferencia}} = \sqrt{\frac{(N_1-1)(S_1^2)+(N_2-1)(S_2^2)}{N_1+N_2-2}\left(\frac{1}{N_1}+\frac{1}{N_2}\right)} \qquad \text{(10-11)}$$

Después se calcula t en la forma usual: $(M_1 - M_2)/S_{\text{Diferencia}}$.

La tabla 10-10 indica los cálculos de $S_{\text{Diferencia}}$ (y t) utilizando la ecuación 10-11, en la comparación realizada por Valenzuela (1997) entre la asistencia instructiva brindada por las madres de niños pobres adecuadamente alimentados y la proporcionada por las madres de niños pobres crónicamente desnutridos. Compare estos cálculos con los de la tabla 10-2.

Tabla 10-10.
Cálculos de la prueba t para medias independientes correspondientes al estudio acerca de la calidad de asistencia por parte de madres de niños pobres chilenos adecuadamente alimentados en comparación con niños pobres chilenos crónicamente desnutridos, utilizando la fórmula de cálculo para $S_{\text{Diferencia}}$.

Niños adecuadamente alimentados

$N_1 = 43$; $gl_1 = N_1 - 1 = 42$; $M_1 = 33{,}1$; $S_1^2 = 201{,}64$

Niños crónicamente desnutridos

$N_2 = 42$; $gl_2 = N_2 - 1 = 41$; $M_2 = 27{,}0$; $S_2^2 = 134{,}56$

$$S_{\text{Diferencia}} = \sqrt{\frac{(N_1-1)(S_1^2)+(N_2-1)(S_2^2)}{N_1+N_2-2}\left(\frac{1}{N_1}+\frac{1}{N_2}\right)}$$

$$= \sqrt{\frac{(43-1)(201{,}64)+(42-1)(134{,}56)}{43+42-2}\left(\frac{1}{43}+\frac{1}{42}\right)}$$

$$= \sqrt{\frac{(42)(201{,}64)+(41)(134{,}56)}{83}(0{,}023+0{,}024)}$$

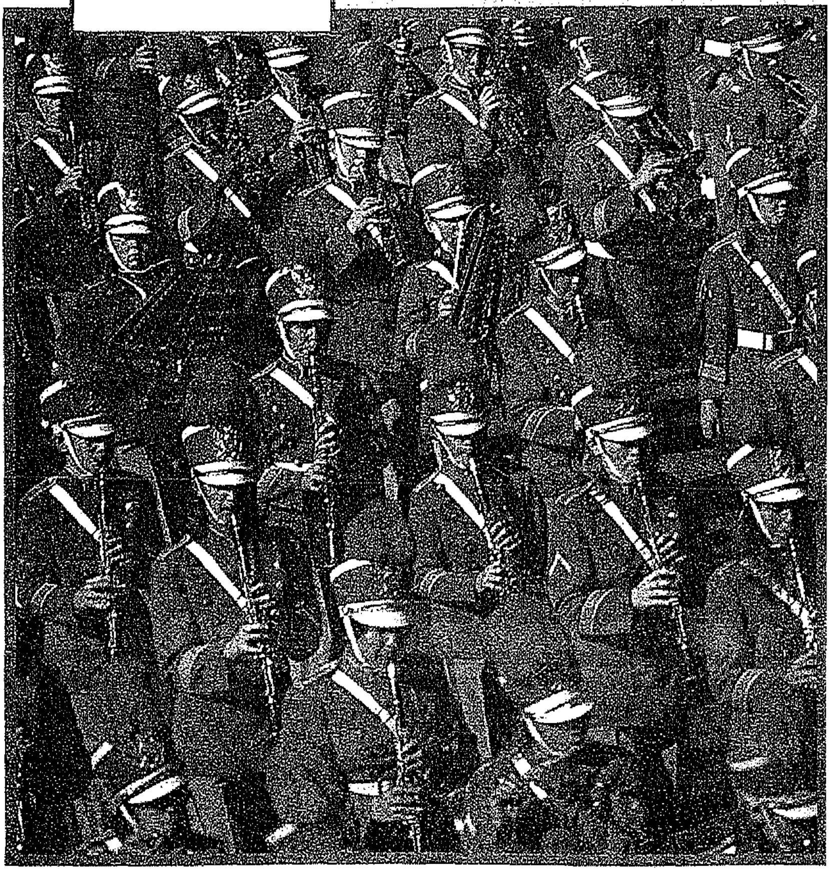
$$= \sqrt{\frac{8.468{,}88+5.516{,}96}{83}(0{,}047)}$$

$$= \sqrt{\frac{13.985{,}84}{83}(0{,}047)} \quad = \sqrt{(168{,}50)(0{,}047)} = \sqrt{7{,}92} = 2{,}81$$

t necesario para nivel 5%, $gl = 83$ (utilizando $gl = 80$ de la tabla), y prueba de dos colas $= 1{,}990$.
$t = (M_1 - M_2)/S_{\text{Diferencia}} = (33{,}1 - 27{,}0)/2{,}81 = 6{,}1/2{,}81 = 2{,}17$.
Conclusión: se rechaza la hipótesis nula y se sostiene la hipótesis de investigación.

Fuente: Valenzuela (1997).

11 Introducción al análisis de la varianza



Descripción del capítulo

Cindy Hazan y Philip Shaver (1987) realizaron los arreglos necesarios para que el *Rocky Mountain News*, un importante periódico de la zona de Denver, imprimiera una encuesta que se distribuiría con el diario. La encuesta incluía la pregunta que aparece en la tabla 11-1, cuya finalidad es realizar una medición de estilos de vinculación. Aquellos que eligieron la primera opción son individuos "seguros"; los que eligieron la segunda, "evasivos", y los que eligieron la tercera, "ansiosos-ambivalentes". Los estilos de vinculación mencionados se consideran formas diferentes de actuar y pensar en lo que respecta a las relaciones personales estrechas, formas que se desarrollan a partir de la experiencia de cada persona con quienes se hicieron cargo de cuidarlos desde temprana edad (p. ej. Mickelson et al., 1997). Los lectores también respondieron preguntas acerca de una cantidad de aspectos relacionados con el amor, entre los que se incluía el nivel de celos que experimentaran. Posteriormente, Hazan y Shaver compararon el nivel de celos registrado para personas con diferentes estilos de vinculación.

Con una prueba *t*, Hazan y Shaver podrían haber comparado las medias de los valores del nivel de celos entre dos de los estilos de vinculación. Pero, en realidad, estaban interesados en las diferencias entre los tres estilos de vinculación. El procedimiento estadístico para probar diferencias entre medias de varios grupos se denomina ANOVA (*Analysis of Variance*, Análisis de varianza). (El análisis de varianza se podría utilizar para un estudio con sólo dos grupos, pero la prueba *t*, que en ese caso da los mismos resultados, es más simple).

En este capítulo, presentamos el análisis de varianza concentrándonos en la situación en la que los diferentes grupos comparados tienen la misma cantidad de valores observados. La situación más complicada, en la que la cantidad de personas en cada grupo no es la misma, será tratada en el capítulo 12. En el capítulo 13, completamos el estudio del análisis de varianza analizando situaciones en las que los distintos grupos se organizan a partir de más de una dimensión. Por ejemplo, en el mismo análisis podríamos tener en cuenta tanto el sexo como el estilo de vinculación, con lo cual se crearían seis grupos en total (femenino seguro, masculino seguro, femenino

Tabla 11-1.
Pregunta utilizada en la encuesta realizada por Hazan y Shaver (1987) a través de un periódico.

¿Cuál de las siguientes posibilidades describe mejor sus sentimientos? [Marque una].

[ ] Me resulta relativamente fácil acercarme a los demás y me siento cómodo si confío en ellos y sé que confían en mí. En líneas generales no me preocupo por la posibilidad de ser abandonado o de que alguien se acerque demasiado a mí.

[ ] Estar cerca de otros me hace sentir un poco incómodo; me resulta difícil confiar completamente en los demás y permitirme depender de ellos. Me pone nervioso que alguien se acerque demasiado a mí, y mis parejas a menudo me piden una relación más íntima de la que puedo mantener sintiéndome cómodo.

[ ] Me parece que los demás no quieren acercarse a mí tanto como yo quisiera. Con frecuencia me preocupo porque pienso que mi pareja realmente no me ama o no va a querer permanecer a mi lado. Quiero unirme completamente con otra persona, y este deseo a veces las ahuyenta.

Fuente: Hazan & Shaver (1987), p. 515.

ansioso, etc.), formados conforme a las dos dimensiones: sexo y estilo de vinculación. El caso que acabamos de describir se conoce como "análisis factorial de varianza". Para acentuar la diferencia con el análisis factorial de varianza, lo que aprenderemos en este capítulo y el siguiente se denomina con frecuencia **análisis de varianza de un criterio**. (No debemos preocuparnos ahora si el concepto de dimensiones resulta confuso. Ya lo trataremos pausada y sistemáticamente en el capítulo 13; sólo lo mencionamos ahora para que el alumno no se sorprenda si llegara a encontrarse con esos términos).

LÓGICA BÁSICA DEL ANÁLISIS DE VARIANZA

La hipótesis nula en un análisis de varianza establece que las diversas poblaciones que se comparan tienen la misma media. Por ejemplo, en el estudio de Hazan y Shaver la hipótesis nula establecería que las poblaciones de personas seguras, ansiosas y evasivas presentan todas el mismo nivel de celos, es decir, que la media en cuanto a celos es la misma en las tres poblaciones. La hipótesis de investigación establecería que el nivel de celos difiere entre las tres poblaciones, es decir, que sus medias no son todas iguales.

La prueba de hipótesis con análisis de varianza trata de probar si las medias muestrales difieren más de lo que esperaríamos si la hipótesis nula fuera verdadera. Sorprendentemente, esta cuestión sobre **medias** se responde analizando **varianzas** (por eso el nombre **análisis de varianza**). (Para expresarlo de modo más simple, nos concentramos en las varianzas porque cuando estamos interesados en el grado en el que difieren varias medias entre sí, lo que estamos estudiando es la variación entre esas medias).

Por lo tanto, para comprender la lógica del análisis de varianza nos dedicaremos a estudiarlas. Particularmente, comenzamos analizando dos formas diferentes de estimar las varianzas poblacionales. Como veremos, el análisis de varianza es una comparación de los resultados de estas dos maneras diferentes de estimar las varianzas de la población.

Estimación de la varianza poblacional a partir de la variación dentro de cada muestra

En el análisis de varianza, como en la prueba t, no conocemos las verdaderas varianzas poblacionales, pero al igual que con la prueba t, la varianza de las poblaciones puede estimarse a partir de los valores de cada muestra. También, al igual que con la prueba t, en el análisis de varianza suponemos que todas las poblaciones tienen la **misma** varianza. Como se supone que todas las poblaciones tienen la misma varianza, las estimaciones realizadas a partir de los valores de cada muestra pueden combinarse o promediarse para formar una sola y mejor estimación. La estimación combinada resultante se denomina **estimación intragrupal de la varianza poblacional**. Lleva este nombre porque es un promedio de las estimaciones calculadas completamente a partir de los valores **dentro** de cada muestra.

Lo más importante que debemos recordar acerca de esta estimación intragrupal es que no se ve afectada por el hecho de que la hipótesis nula sea verdadera. Es decir, la estimación será la misma ya sea porque las medias poblacionales son todas iguales (como lo serían si la hipótesis nula fuera verdadera) o porque las medias poblacionales son diferentes (como lo serían si la hipótesis nula no fuera verdadera). La estimación será la misma porque sólo se concentra en la variación dentro de cada población, y no importa cuán alejadas estén las medias de las distintas poblaciones.

Estimación de la varianza poblacional a partir de la variación entre las medias muestrales

En esta sección veremos el otro modo de estimar la varianza poblacional. La media de cada muestra es un número por sí mismo. Si existen varias muestras, hay varios números, y estos números tendrán una verdadera variación entre sí. Sucede que la variación entre esas medias nos ofrece otro modo de estimar la varianza en las poblaciones de las que provienen las muestras. La forma en que esto funciona es un poco intrincada, por lo que recomendamos prestar mucha atención a la siguiente sección.

Cuando la hipótesis nula es verdadera. En primer lugar analizaremos la situación en la que la hipótesis nula es verdadera, de modo que todas las muestras provienen de poblaciones con la misma media. No debemos olvidar que estamos suponiendo que todas las poblaciones tienen la misma varianza (y además todas son normales). Por lo tanto, si la hipótesis nula es verdadera, todas las poblaciones son idénticas (tienen la misma media, varianza y forma).

Si trabajamos con muestras de varias poblaciones idénticas, aun cuando las poblaciones sean idénticas, las muestras serán levemente diferentes entre sí, y sus medias también lo serán. ¿En qué medida pueden ser diferentes las medias? Eso depende de cuánta variación haya dentro de cada población. Si una población tiene muy poca variación entre los valores que la conforman, entonces las medias de las muestras de esa población tenderán a ser muy similares entre sí. Lo mismo ocurre con las medias de varias poblaciones idénticas; la media de la muestra de una de las poblaciones tenderá a ser similar a la media de una muestra de cualquier otra.

¿Qué sucede si varias poblaciones idénticas presentan mucha variación entre los valores que las conforman? En ese caso, si tomamos una muestra de cada población, las medias de esas muestras podrían fácilmente ser muy diferentes entre sí y, al serlo, habrá entre ellas una gran varianza. El tema es que cuanta más varianza haya dentro de cada una de varias poblaciones idénticas, mayor varianza habrá entre las medias muestrales cuando se seleccione una muestra de cada población.

Veamos un ejemplo. Supongamos que estuviéramos estudiando muestras formadas por seis niños de cada una de las tres grandes clases (las poblaciones del ejemplo). Si cada clase estuviera formada por niños que tuvieran 9 ó 10 años de edad, las medias de las tres muestras serían valores entre 9 y 10, es decir que no habría mucha varianza entre esas medias. Pero si cada clase estuviera formada por niños de 5 a 15 años de edad, las medias de las tres muestras probablemente variarían bastante. Es decir, la variación entre las medias muestrales está directamente relacionada con el grado de variación dentro de cada una de las poblaciones de las que provienen las muestras. A mayor variación en cada población, mayor variación entre las medias de las muestras tomadas de esas poblaciones.

Analicemos el ejemplo de las poblaciones de estilos de vinculación seguro, ansioso y evasivo estudiado por Hazan y Shaver. Por supuesto que habrá alguna varianza en el nivel de celos de las diferentes personas dentro de cada una de esas poblaciones. Pero supongamos por ahora que las tres poblaciones tienen todas la misma media en cuanto al nivel de celos (como sucedería si la

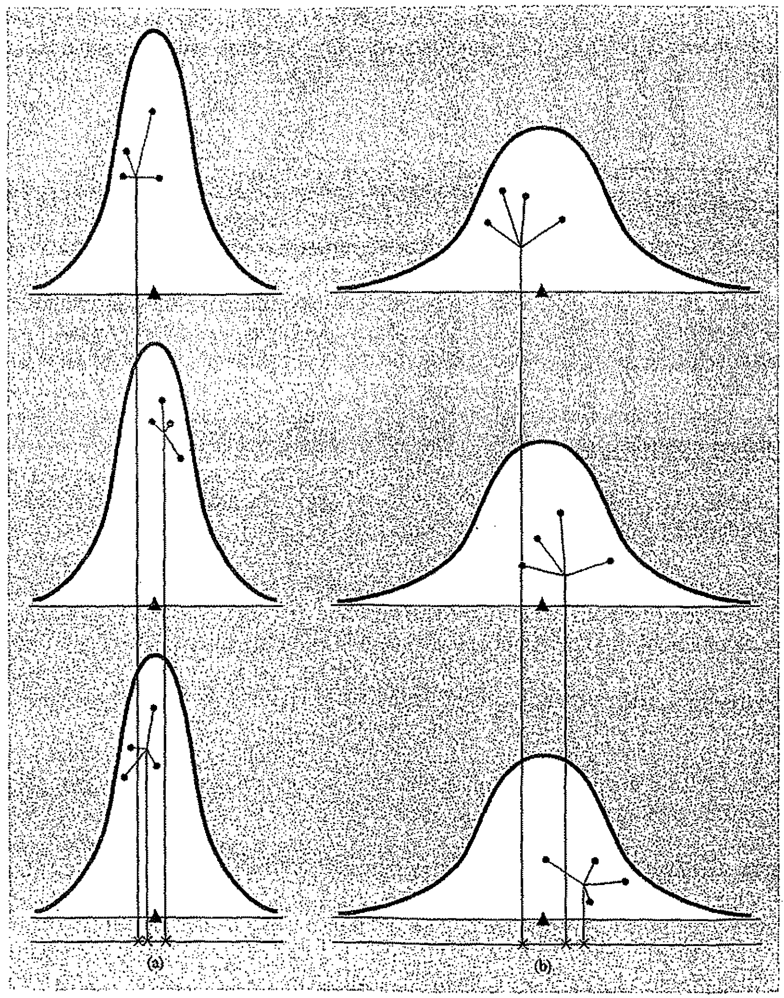


Figura 11-1. Las medias de muestras provenientes de poblaciones idénticas no serán idénticas. Las medias muestrales que provienen de poblaciones con menos variación, variarán menos (a). Las medias muestrales que provienen de poblaciones con más variación, variarán más (b). Las medias poblacionales se indican con un triángulo; las medias muestrales con una X.

hipótesis nula fuera verdadera). Aun en el caso en que todas tuvieran la misma media, es probable que la muestra de una población no tenga exactamente la misma media que la muestra de una segunda población. Del mismo modo, es probable que la muestra de una tercera población sea levemente diferente de las de las otras dos. Y así sucesivamente. Más aún, cuánto más varíe cada una de esas poblaciones internamente, más variarán las medias de muestras tomadas de esas poblaciones. Variarían incluso si, de hecho, las medias poblacionales fueran idénticas.

La figura 11-1 representa gráficamente el principio que hemos estado analizando. Las tres poblaciones idénticas de la izquierda tienen poca varianza, y las tres poblaciones idénticas de la derecha registran una gran varianza. En cada serie de tres poblaciones idénticas, aun cuando las medias de las tres poblaciones sean iguales, las medias de las muestras provenientes de esas poblaciones no son iguales. Es muy importante destacar que las medias de las poblaciones con menos varianza son más cercanas (tienen menos varianza entre sí) y que las medias de poblaciones con más varianza están más dispersas (tienen más varianza entre sí).

Ya hemos visto que la variación entre las medias de muestras tomadas de poblaciones idénticas está directamente relacionada con la variación de los valores dentro de cada una de esas poblaciones. Esto tiene una implicancia muy importante: sería posible estimar la varianza dentro de cada población a partir de la variación entre las medias de las muestras. Es decir, podríamos utilizar la variación de las medias muestrales para calcular el grado de variación en la población de donde provienen esas muestras.

Tal estimación se denomina **estimación intergrupal de la varianza poblacional.** (Lleva este nombre porque se basa en la variación entre las medias de las muestras, es decir, de los "grupos"). Más adelante, en éste capítulo, veremos cómo se calcula realmente esta estimación.

Hasta aquí, la lógica que hemos analizado supone que la hipótesis nula es verdadera, en cuyo caso no existe variación entre las medias poblacionales. Veamos ahora qué sucede cuando la hipótesis nula no es verdadera y sí lo es la hipótesis de investigación.

Cuando la hipótesis nula no es verdadera. Si la hipótesis nula no es verdadera y la hipótesis de investigación sí lo es, las propias poblaciones tendrán diferentes medias. En ese caso, la variación entre las medias de muestras tomadas de esas poblaciones sigue siendo el resultado de la variación dentro de las poblaciones. La diferencia radica en que, en este caso, en el que la hipótesis de investigación es verdadera, la variación entre medias muestrales es causada **además** por la variación entre las medias poblacionales. Es decir, en este caso las medias muestrales se dispersan por dos razones diferentes: a) por la variación dentro de cada una de las poblaciones y b) por la variación entre las poblaciones. La figura 11-2a representa gráficamente tres poblaciones con las mismas medias y las medias muestrales provenientes de ellas (es decir, la misma situación que en la figura 11-1, a y b). La figura 11-2b representa gráficamente tres poblaciones con diferentes medias y las medias de las muestras tomadas de ellas (es decir, la situación que acabamos de explicar). Vale la pena observar que las medias de las muestras están más dispersas en la figura 11-2b que en la figura 11-2a, aun cuando las variaciones dentro de las poblaciones sean las mismas en 11-2b y en 11-2a. Esta dispersión adicional (varianza) que representa la figura 11-2b se debe a que las poblaciones tienen diferentes medias.

En resumen, la estimación intergrupal de la varianza poblacional se calcula sobre la base de la variación entre las medias muestrales. Si la hipótesis nula es verdadera, esa estimación es una indicación precisa de la variación dentro de las poblaciones. Pero si la hipótesis nula es falsa, este método de estimación de la varianza poblacional se ve influenciado tanto por la variación dentro de las poblaciones como por la variación entre ellas. Por lo tanto, no proporcionará una estimación precisa de la variación dentro de las poblaciones porque también estará afectada por la variación entre las poblaciones. La diferencia que acabamos de mencionar tiene implicancias importantes: es lo que hace del análisis de varianza un método de prueba de hipótesis basado en la existencia o no de diferencias entre las medias de diferentes grupos.

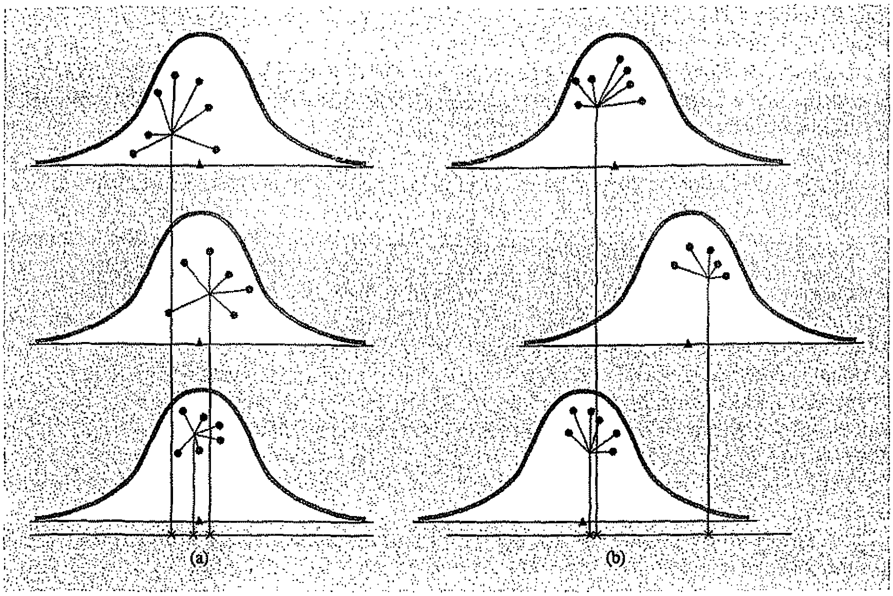


Figura 11-2. Las medias muestrales que provienen de poblaciones cuyas medias son diferentes (b) variarán más que las medias muestrales que provienen de poblaciones cuyas medias son iguales (a). Las medias poblacionales se indican con un triángulo; las medias muestrales con una X.

Comparación de las estimaciones intragrupal e intergrupal de la varianza poblacional

La tabla 11-2 resume lo que hemos visto hasta ahora con respecto a las estimaciones intragrupal e intergrupal de la varianza poblacional, tanto cuando la hipótesis nula es verdadera como cuando también lo es la hipótesis de investigación. Cuando la hipótesis nula es verdadera, las estimaciones intragrupal e intergrupal se basan en lo mismo; literalmente, son estimaciones de la misma varianza poblacional. Cuando la hipótesis nula es verdadera, ambas estimaciones deberían ser aproximadamente iguales (sólo **aproximadamente** iguales, ya que estamos hablando de estimaciones). Otro modo de describir la similitud entre la estimación intergrupal y la estimación intragrupal, cuando la hipótesis nula es verdadera, es la siguiente: la razón entre estimación intergrupal y la estimación intragrupal debería ser aproximadamente 1. Por ejemplo, si la estimación intragrupal es 107,5, la estimación intergrupal debería ser de alrededor de 107,5, de forma tal que la razón fuera aproximadamente 1 (la razón se calcula dividiendo uno de los números por el otro).

De todos modos, la situación es bastante diferente cuando la hipótesis nula no es verdadera. Como lo indica la tabla 11-2, cuando la hipótesis de investigación es verdadera, la estimación intergrupal es influenciada por dos fuentes de variación: a) la variación de los valores dentro de cada población y b) la variación de las medias de las poblaciones entre sí. Sin embargo, incluso cuando la hipótesis de investigación es verdadera, la estimación intragrupal continúa siendo influenciada sólo por la variación dentro de las poblaciones. Es decir, que cuando la hipótesis de investigación es verdadera, la estimación intergrupal debería ser mayor. En este caso, la razón de la

Tabla 11-2.
Fuentes de variación de las estimaciones intragrupal e intergrupal de la varianza.

	Variación dentro de las poblaciones	Variación entre las poblaciones
La hipótesis nula es verdadera		
Estimación intragrupal causada por	X	
Estimación intergrupal causada por	X	
La hipótesis de investigación es verdadera		
Estimación intragrupal causada por	X	
Estimación intergrupal causada por	X	X

estimación intergrupal y la estimación intragrupal debería ser mayor que 1. Por ejemplo, la estimación intergrupal podría ser 638,9 y la estimación intragrupal 107,5, dando una razón de 638,9/107,5, ó lo que es igual, de 5,94. Es decir, si dividimos la estimación mayor, la intergrupal, por la menor, la intragrupal, no obtenemos 1, sino un número mayor.

Lo que acabamos de describir es el principio fundamental del análisis de varianza. Cuando la hipótesis nula es verdadera, la razón entre la estimación de varianza intergrupal y la estimación de varianza intragrupal debería ser aproximadamente 1, pero cuando la hipótesis de investigación es verdadera, la razón debería ser mayor a 1. Por lo tanto, si calculamos la razón y ésta resulta mucho mayor a 1, podemos rechazar la hipótesis nula. Es decir, no es verosímil que la hipótesis nula pueda ser verdadera ya que la estimación intergrupal es mucho mayor que la estimación intragrupal.

La razón *F*

Esta razón fundamental entre las estimaciones intergrupal e intragrupal de la varianza se denomina **razón *F***. (La *F* se debe a Sir Ronald Fisher, un destacado estadístico que desarrolló el análisis de varianza; *véase* cuadro 11-1).

La distribución *F* y la tabla *F*

Ya hemos mencionado que cuando la razón fundamental entre la estimación intergrupal y la estimación intragrupal (la razón *F*) es mucho mayor a 1, podemos rechazar la hipótesis nula. La siguiente pregunta es: ¿Cuánto mayor a 1 necesita ser la razón para que podamos rechazar la hipótesis nula con confianza?

Conforme a lo que el alumno seguramente ya debe haber imaginado, los estadísticos han desarrollado los cálculos matemáticos de una **distribución *F*** y han preparado tablas de razones *F*. Para cualquier situación determinada, simplemente buscamos en una **tabla *F*** cuán extremo debe ser una razón *F* para rechazar la hipótesis nula a, digamos, un nivel 0,05. (*Más adelante*, en este capítulo, aprenderemos a utilizar la tabla *F*).

Para dar un ejemplo de la razón *F*, volvamos al estudio acerca del estilo de vinculación realizado por Hazan y Shaver (1987). Los resultados de ese estudio, en cuanto al nivel de celos, fueron los siguientes: la estimación de varianza poblacional intergrupal era de 23,19. (Este número se calcula sobre la base de las medias de las tres muestras de estilo de vinculación, que eran 2,17, 2,88 y 2,57; pronto aprenderemos a realizar estos cálculos). *La estimación intragrupal* de varianza poblacional era 0,53. (Este número se calculó combinando las estimaciones de la varianza de

Cuadro 11-1.
Sir Ronald Fisher, genio mordaz de la estadística.

Ronald A. Fisher, contemporáneo de William Gosset (*véase* cuadro 9-1) y de Karl Pearson (*véase* cuadro 14-1), fue probablemente el más brillante y productivo de los miembros del cerrado grupo de estadísticos británicos. A lo largo del proceso de elaboración de trescientos trabajos y de siete libros, desarrolló muchos de los conceptos clave de la ciencia moderna: varianza, análisis de varianza, estadísticas (en el sentido de describir una muestra, en oposición con los parámetros de una población), niveles de significación, hipótesis nula y casi también de todas las ideas básicas del diseño de investigación, además de señalar la importancia fundamental de la aleatorización.

Una de las tantas leyendas familiares cuenta que el pequeño Ronald, nacido en el año 1890 en East Finchley, un suburbio del norte de Londres, estaba tan fascinado por la matemática que, un día, a los 3 años de edad, al ser puesto frente a su silla alta para el desayuno, le preguntó a su niñera: "¿Qué es la mitad de la mitad?" Cuando se le explicó que era un cuarto, preguntó "¿Qué es la mitad de un cuarto?" Al recibir la respuesta quiso saber qué era la mitad de un octavo. Ante la siguiente respuesta supuestamente pensó un momento y dijo "Entonces supongo que la mitad de un dieciseisavo debe ser un treintaidosavo". En fin, historias de niños.

Sin embargo, cuando Fisher llegó a la adultez, parece haber estado muy lejos de ser adorable. Algunos observadores atribuyen esta característica al hecho de que tuvo una madre fría y poco emotiva. Cualquiera sea la razón, durante su vida, Fisher se vio involucrado en profundas enemistades, incluso con alumnos que previamente habían sido sus más cercanos aliados y quienes verdaderamente deberían haber sido sus compañeros de investigación. Cuando le hablaban en broma él contestaba con seriedad mortal; cuando los demás estaban serios él bromeaba. En una oportunidad, relata William G. Cochran (otro estadístico reconocido), estaba por cruzar una calle junto con Fisher. El momento no era el más indicado y , ante la vacilación de Cochran, Fisher lo increpó: "¡Ah, vamos!, ¡no nos lastimará un poco de selección natural"!, y Cochran tímidamente arriesgó su vida.

La poca compasión de Fisher se extendía también a sus lectores: su estilo no sólo era terriblemente oscuro sino que, con frecuencia, omitía explicitar importantes supuestos y pruebas. Gosset mismo expresó que cuando Fisher comenzaba una oración con **evidentemente**, eso significaba dos horas de arduo trabajo antes de que uno pudiera tener esperanzas de discernir por qué el tema era evidente. Otro estadístico buscó excusarlo, sin embargo, diciendo que "Fisher hablaba en un nivel escasamente comprendido por el resto de la humanidad". Y es verdad que era invariablemente admirado y respetado por su trabajo, aunque no por sus modales.

De hecho, su falta de empatía se extendía a toda la humanidad. Al igual que Galton, Fisher estaba a favor de la eugenesia; favorecía todo aquello que pudiera aumentar el índice de natalidad de las clases altas y profesionales, como también de los artesanos capacitados. Él no sólo pensaba que la anticoncepción era una mala idea –temía que las personas cuya descendencia era menos deseable recurrieran a ella en menor proporción–, sino que defendía el infanticidio como herramienta de la función evolutiva. Probablemente también haya sido un acto de justicia que sus oportunidades de experimentar con la reproducción nunca hayan llegado más allá de la crianza de sus propios hijos y de algunos cultivos de papa y trigo.

Lo que con más fuerza influyó en Fisher fueron probablemente sus catorce años de trabajo en una estación experimental agrícola llamada Rothamsted, en Hertfordshire, 25 millas al norte de Londres. En Rothamsted, Fisher, al igual que Gosset en su fábrica de cerveza en Dublín, enfrentó todo tipo de problemas prácticos, tales como averiguar si las aplicaciones anuales de abono mejoraban el rendimiento del campo a largo plazo o si eran la causa de misteriosas disminuciones de producción luego de muchas décadas. Tal vez fue este aislamiento de las disputas personales entre los académicos de Londres y la cercanía a los temas reales los que ayudaron a Fisher a concentrarse en el desarrollo de la estadística como una poderosa herramienta metodológica.

Aunque con el tiempo Fisher accedió al cargo de titular de la cátedra de Eugenesia en el University College, su oportunidad profesional de mayor influencia tal vez hay sido la invitación a la Facultad del Estado de Iowa, en Ames, en los veranos de 1931 y 1936 (donde, según se dice, estaba muy perturbado por el terrible calor que guardaban sus sábanas todo el día en el refrigerador). En Ames, Fisher provocó una fuerte impresión en George Snedecor, un profesor de matemática estadounidense que también investigaba problemas agrícolas. Posteriormente, Snedecor escribió un libro sobre estadística aplicada a la agricultura, que tomaba muchas ideas del trabajo de Fisher en Rothamsted. El libro difundió a tal punto las ideas de Fisher sobre estadística y diseño de investigación, que su segunda edición vendió 100.000 copias.

Durante su estadía en Ames, Fisher también se ganó la admiración de E. F. Lindquist, profesor de educación en la Universidad de Iowa, con sede en la misma ciudad. El siguiente libro de Lindquist estuvo totalmente permeado con las ideas de Fisher aplicadas al campo de la educación y la psicología, áreas en las que han desempeñado un papel primordial hasta la actualidad.

cada población sobre la base de los valores de cada muestra). La razón entre las estimaciones de varianza intergrupal e intragrupal (23,19/0,53) resulta ser 43,91; es decir $F = 43{,}91$. La razón F calculada es considerablemente mayor a 1. De hecho, la razón F necesaria para rechazar la hipótesis nula al nivel 0,05 es sólo 3,01. Hazan y Shaver rechazaron con confianza la hipótesis nula, y concluyeron que el nivel de celos varía según el estilo de vinculación.

Una analogía

Para algunos estudiantes, la siguiente analogía les resulta de gran ayuda para comprender el análisis de varianza. La analogía se realiza con lo que los ingenieros llaman razón señal-ruido. Por ejemplo, la capacidad de comprender las palabras en una conversación por teléfono celular con interferencia depende de la potencia de la señal, en contraposición con la cantidad de ruido aleatorio. En el caso de la razón F en el análisis de varianza, la diferencia entre las medias de las muestras se equipara con la señal, es la información de interés, y la variación dentro de las muestras se equipara con el ruido. Cuando la variación entre las muestras es lo suficientemente grande en comparación con la variación dentro de las muestras, la conclusión es que existe un efecto significativo.

REALIZACIÓN DE UN ANÁLISIS DE VARIANZA

Luego de haber estudiado la lógica básica del análisis de varianza, analizaremos un ejemplo para ilustrar los detalles. (Utilizamos un estudio ficticio para que los números sean simples).

Supongamos que un psicólogo especializado en temas sociales está estudiando la influencia del conocimiento de la existencia de antecedentes criminales en la percepción del jurado con respecto a la culpabilidad o inocencia del acusado. El investigador recluta 15 voluntarios que han sido seleccionados para integrar un jurado (pero que todavía no han actuado en un juicio). El investigador les muestra una filmación de video de un juicio de cuatro horas de duración en el que una mujer es acusada de entregar cheques falsos. Antes de ver la cinta, se entrega a todos los participantes una "hoja de antecedentes" con la edad, estado civil, educación y otros datos sobre la acusada. La hoja es la misma para los 15 participantes, con una diferencia: en el caso de 5 de los participantes, la última sección de la hoja dice que la mujer ha sido condenada varias veces antes de entregar los cheques falsos. (Llamaremos a los participantes que recibieron esta versión de la hoja de antecedentes "grupo del informe delictivo"). En el caso de otros 5 de los participantes, la última sección de la hoja dice que la mujer tiene una historia delictiva completamente limpia ("grupo del informe en blanco"). Finalmente, en el caso de los cinco participantes restantes, la hoja no hace ninguna mención acerca de antecedentes delictivos ("grupo sin información").

Los participantes son asignados a los grupos al azar. Después de ver los videos del juicio, los 15 participantes califican a la acusada con una escala de 10 puntos, que va desde "completamente seguro de que es inocente" (1) a "completamente seguro de que es culpable" (10). Los resultados del estudio ficticio se indican en la tabla 11-3. La tabla muestra que las medias de los tres grupos son diferentes (8, 4 y 5), pero que además hay bastante variación dentro de cada uno de los tres grupos (las estimaciones de varianza poblacional realizadas a partir de los valores de estos tres grupos son 4,5, 5,0 y 6,5).

Necesitamos realizar tres cálculos para probar la hipótesis que establece que las tres poblaciones son diferentes: a) una estimación de varianza poblacional sobre la base de la variación de los valores dentro de cada muestra; b) una estimación de la varianza poblacional sobre la base de las diferencias entre las medias de los grupos, y c) la razón de las dos, es decir, la razón *F*. (Además, necesitamos el punto de corte correspondiente al nivel de significación elegido tomado de una tabla *F*). Analicemos cada uno de estos cálculos por vez.

Tabla 11-3.
Resultados del estudio acerca de antecedentes delictivos. (Datos ficticios).

	Grupo del informe delictivo			Grupo del informe en blanco			Grupo sin información		
	Calificación	*Desvío de la media*	*Desvío cuadrático de la media*	*Calificación*	*Desvío de la media*	*Desvío cuadrático de la media*	*Calificación*	*Desvío de la media*	*Desvío cuadrático de la media*
	10	2	4	5	1	1	4	–1	1
	7	–1	1	1	–3	9	6	1	1
	5	–3	9	3	–1	1	9	4	16
	10	2	4	7	3	9	3	–2	4
	8	0	0	4	0	0	3	–2	4
Σ:	40	0	18	20	0	20	25	0	26

$M = 40/5 = 8$.
$S^2 = 18/4 = 4{,}5$

$M = 20/5 = 4$
$S^2 = 20/4 = 5{,}0$

$M = 25/5 = 5$
$S^2 = 26/4 = 6{,}5$

Estimación de la varianza poblacional a partir de la variación entre los valores dentro de cada grupo

La varianza poblacional puede estimarse a partir de cualquiera de los grupos (es decir, a partir de cualquier muestra) utilizando el método usual para estimar la varianza poblacional a partir de una muestra. Primero, calculamos la suma de los desvíos cuadráticos, es decir, tomamos el desvío de cada registro con respecto a la media de su grupo, elevamos el desvío al cuadrado y sumamos todos los desvíos cuadráticos. Segundo, dividimos esa suma de desvíos cuadráticos por los grados de libertad del grupo (los grados de libertad de un grupo constituyen la cantidad de valores observados en el grupo menos 1). En el ejemplo, como lo indica la tabla 11-3, esto da una varianza poblacional estimada de 4,5 sobre la base del grupo del informe delictivo, una estimación de 5,0 sobre la base del grupo del informe en blanco, y una estimación de 6,5 sobre la base del grupo sin información.

No debemos olvidar que en el análisis de varianza, al igual que en la prueba *t*, se supone que las poblaciones tienen la misma varianza. Dado que estas estimaciones pertenecen a poblaciones que se supone tienen la misma varianza, las estimaciones basadas en los valores de cada muestra están estimando todas el mismo número (la verdadera varianza poblacional). Además, dado que los tamaños de muestra en este ejemplo son iguales, cada grupo representa una estimación basada en la misma cantidad de información; por lo tanto, podemos combinar estas estimaciones de varianza realizando un promedio ordinario. El resultado es una estimación general de la varianza poblacional sobre la base de la variación dentro de los grupos, que es igual a la suma de 4,5, 5,0 y 6,5 (o sea 16) dividida por la cantidad de grupos (o sea 3). El resultado es 5,33.

La varianza estimada sobre la base de la variación de los valores dentro de cada uno de los grupos es la estimación intragrupal de la varianza. Se simboliza como S^2_{Dentro} ó CM_{Dentro}. CM_{Dentro} es la abreviatura de **cudrado medio dentro**. El término cuadrado **medio** de los cuadrados es otro nombre de la varianza, ya que la varianza es la media de los desvíos cuadráticos. (La S^2_{Dentro} ó CM_{Dentro} también se denomina a veces "varianza del error", y se simboliza como S^2_{Error} ó CM_{Error}.)

La fórmula para la estimación intragrupal de varianza, cuando los tamaños de las muestras son iguales, es:

$$S^2_{\text{Dentro}} \text{ ó } CM_{\text{Dentro}} = \frac{S_1^2 + S_2^2 + \cdots + S^2_{\text{Último}}}{N_{\text{Grupos}}} \tag{11-1}$$

En la fórmula, S_1^2 es la varianza poblacional estimada sobre la base de los valores del primer grupo (el que proviene de la población 1); S_2^2 es la varianza poblacional estimada sobre la base de los valores del segundo grupo; $S^2_{\text{Último}}$ es la varianza poblacional estimada sobre la base de los valores del último grupo. (Los puntos, o elipsis, en la fórmula indican que debemos completarla con la varianza poblacional estimada correspondiente a todos los otros grupos que hay en el análisis). N_{Grupos} es la cantidad grupos.

Utilizando esta fórmula para realizar los cálculos, obtenemos:

$$S^2_{\text{Dentro}} = \frac{S_1^2 + S_2^2 + \cdots + S^2_{\text{Último}}}{N_{\text{Grupos}}} = \frac{4{,}5 + 5{,}0 + 6{,}5}{3} = \frac{16{,}3}{3} = 5{,}33$$

Estimación de la varianza poblacional a partir de las diferencias entre las medias de los grupos

Determinar la estimación intergrupal de la varianza poblacional involucra dos pasos: primero, estimar a partir de unas pocas medias (las medias muestrales) la varianza de una distribución de medias (la distribución de todas las medias muestrales posibles de una población de observaciones individuales). Segundo, sobre la base de la varianza de esa distribución de medias, calcular la varianza poblacional de observaciones individuales.

Estimación de la varianza de la distribución de medias. Podemos considerar que las medias muestrales fueron tomadas de una distribución de medias muestrales. Seguimos el procedimiento usual utilizando los valores de una muestra para estimar la varianza de la población de donde provienen esos valores. En este caso, consideramos las medias muestrales como los valores y la distribución de medias como la población de donde provienen esos valores. Todo esto queda reducido a los siguientes pasos: comenzamos calculando la suma de los desvíos cuadráticos (calculamos la media de las medias muestrales, calculamos el desvío de cada media muestral con respecto a la media de medias, elevamos al cuadrado cada uno de esos desvíos y, después sumamos los desvíos cuadráticos). Luego, dividimos esa suma de desvíos cuadráticos por los grados de libertad, que es la cantidad de medias menos 1.

Lo anterior se expresa bajo la siguiente fórmula (cuando los tamaños de las muestras son iguales),

$$S_M^2 = \frac{\Sigma(M - GM)^2}{gl_{\text{Entre}}} \tag{11-2}$$

En la fórmula mencionada arriba, S_M^2 es la varianza estimada de la distribución de medias (estimación basada en las medias de las muestras del estudio). M es la media de cada una de las muestras. GM es la **gran media**, la media general de todos los valores, que es también la media de las medias. gl_{Entre} son los grados de libertad en la estimación intergrupal, la cantidad de grupos menos 1. Lo anterior se expresa bajo la fórmula,

$$gl_{\text{Entre}} = N_{\text{Grupos}} - 1 \tag{11-3}$$

En el ejemplo referido a los antecedentes delictivos, las tres medias son 8, 4 y 5. Los cálculos aparecen en la tabla 11-4.

De la varianza estimada de la distribución de medias a una varianza estimada de la población de valores observados. Lo que acabamos de calcular a partir de una muestra de unas pocas medias es la varianza estimada de una distribución de medias. A partir de ese dato queremos estimar la varianza de la población (la distribución de valores individuales) en la que se basa esa distribución de medias. En el capítulo 7 vimos que la varianza de una distribución de medias es menor que la varianza de la población en la que se basa (la distribución de valores individuales). Esto ocurre porque las medias tienen menos posibilidades de ser extremas que los valores individuales (ya que es poco probable que varios valores extremos en la misma dirección puedan quedar incluidos en una misma muestra). Específicamente, en el capítulo 7 aprendimos que la varianza de una distribución de medias es la varianza de la distribución de valores individuales dividida por la cantidad de valores de cada muestra.

Tabla 11-4.
Varianza de la distribución de medias, estimada sobre la base de las medias de los tres grupos experimentales del estudio referido a los antecedentes delictivos (datos ficticios).

	Medias muestrales	Desvíos de la gran media	Desvío cuadrático de la gran media
	(M)	$(M - GM)$	$(M - GM)^2$
	4	–1,67	2,79
	8	2,33	5,43
	5	–0,67	0,45
Σ:	17	-0,01	8,67

$GM = \Sigma M/N_{\text{Grupos}} = 17/3 = 5{,}67$; $S_M^2 = \Sigma(M - GM)^2/gl_{\text{Entre}} = 8{,}67/2 = 4{,}34$.

Ahora, sin embargo, vamos a revertir lo que hicimos en el capítulo 7, en el que calculamos la varianza de la distribución de medias **dividiendo** la varianza de la distribución de observaciones individuales por el tamaño de la muestra. Ahora vamos a calcular la varianza de la distribución de valores individuales **multiplicando** la varianza de la distribución de medias por el tamaño de la muestra. (*Véase* tabla 11-5). Es decir, para obtener la varianza poblacional de observaciones individuales, multiplicamos la estimación de la varianza de la distribución de medias por el tamaño de la muestra. El resultado de este proceso es la estimación intergrupal de la varianza. Lo anterior se expresa bajo la fórmula (en los casos en que los tamaños de muestra son iguales),

$$S^2_{\text{Entre}} \text{ ó } CM_{\text{Entre}} = (S^2_M)(n) \tag{11-4}$$

En la fórmula arriba mencionada, S^2_{Entre} ó CM_{Entre} es la estimación de la varianza poblacional sobre la base de la variación entre las medias (la estimación intergrupal de varianza). n es la cantidad de observaciones de cada muestra.

Volvamos al ejemplo referido a los antecedentes delictivos, en el que había 5 individuos en cada muestra y una varianza estimada de distribución de medias de 4,34. En ese ejemplo, multiplicando 4,34 por 5 obtenemos una estimación intergrupal de varianza poblacional igual a 21,7. Se expresa bajo la fórmula,

$$S^2_{\text{Entre}} \text{ ó } CM_{\text{Entre}} = (S^2_M)(n) = (4{,}34)(5) = 21{,}7$$

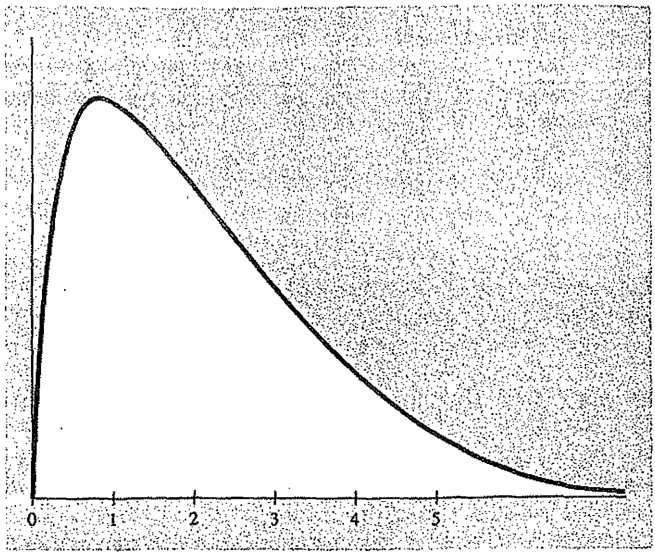
Resumiendo, el procedimiento de estimación de la varianza poblacional, sobre la base de las diferencias entre las medias de los grupos, es el siguiente: a) calcular la varianza estimada de la distribución de medias y luego b) multiplicar esa varianza estimada por la cantidad de observaciones de cada grupo.

Tabla 11-5.
Comparación del cálculo de varianza de una distribución de medias a partir de la varianza de una distribución de observaciones individuales al revés.

De distribución de individuos a distribución de medias: $S^2_M = S^2/N$

De distribución de medias a distribución de individuos: $S^2 = S^2 = (S^2_M)\ (N)$

Figura 11-3.
Una distribución F.



Cálculo de la razón F

La razón F es el cociente entre la estimación intergrupal de la varianza poblacional y la estimación intragrupal de la varianza poblacional. Se representa bajo la fórmula,

$$F = \frac{S^2_{\text{Entre}}}{S^2_{\text{Dentro}}} \text{ ó } \frac{CM_{\text{Entre}}}{CM_{\text{Dentro}}} \tag{11-5}$$

En el ejemplo que analizamos, el ratio entre intergrupal e intragrupal es el cociente entre 21,7 y 5,33. Realizando la división obtenemos una razón F de 4,07. Se expresa bajo la fórmula,

$$F = \frac{S^2_{\text{Entre}}}{S^2_{\text{Dentro}}} \text{ ó } \frac{CM_{\text{Entre}}}{CM_{\text{Dentro}}} = \frac{21{,}7}{5{,}33} = 4{,}07$$

La distribución F

El siguiente paso es determinar el punto de corte a partir del cual se considera que F es lo suficientemente grande como para rechazar la hipótesis nula. Esto requiere una distribución de razones F que podamos utilizar para establecer qué es una razón F extrema.

En la práctica, simplemente buscamos el punto de corte necesario en una tabla. Pero para entender de dónde proviene el número de la tabla, necesitamos comprender la distribución F. La manera más fácil de comprender esta distribución es analizar cómo haríamos para elaborar una.

Supongamos que comenzamos con tres poblaciones idénticas. Después, seleccionamos al azar cinco valores de cada una, y sobre la base de esas tres muestras (de cinco valores cada una), calculamos la razón F. (Es decir, utilizamos estos valores para calcular una estimación intergrupal y una estimación intragrupal, y después dividimos la primera por la segunda). Digamos que al realizar ese proceso, la razón F a la que llegamos es de 1,36. Ahora bien, seleccionamos otras tres muestras al azar de cinco valores cada una y calculamos la razón F utilizando estas tres muestras. Tal vez obtenemos una F de 0,93. Si realizamos todo este proceso muchas veces, finalmente obtendremos muchas razones F. La distribución de todas las razones F posibles calculadas del modo descripto (utilizando muestras aleatorias de poblaciones idénticas) se denomina distribución F. La figura 11-3 muestra un ejemplo de distribución F. (Existen muchas diferentes distribuciones F, y cada una tiene una forma levemente distinta. La forma exacta depende de cuántas muestras tomemos cada vez y de cuántos valores haya en cada muestra, pero la forma general es similar a la que aparece en la figura).

En realidad, nadie se dedica a elaborar su propia distribución F del modo arriba mencionado. Se trata de una distribución matemática cuyas características exactas pueden encontrarse a partir de una fórmula. También puede probarse matemáticamente que si tuviéramos la paciencia de seguir este procedimiento el tiempo necesario, obtendríamos el mismo resultado.

Como podemos observar en la figura 11-3, la distribución F no es simétrica sino que tiene una larga cola hacia la derecha. La razón de esta asimetría positiva es que una distribución F es una distribución de razones de varianzas, y las varianzas siempre son números positivos (una varianza es un promedio de desvíos cuadráticos, y cualquier número elevado al cuadrado es positivo). Una razón entre un número positivo y otro número positivo nunca será menor a 0. Pero nada impide que una razón sea un número muy alto. Por lo tanto, las razones F no pueden ser menores que 0 y pueden ser bastante altas.[1] (La mayoría de las razones F se apilan cerca del valor 1, pero se dispersan más sobre el lado positivo, donde tienen espacio para dispersarse).

La tabla F

La tabla F es un poco más complicada que la tabla t, ya que existe una distribución F diferente según los grados de libertad utilizados en la estimación intergrupal de varianza y según los grados de libertad utilizados en la estimación intragrupal de varianza. Es decir, deben considerarse dos tipos distintos de grados de libertad para buscar el punto de corte necesario. Unos son los **grados de libertad entre**, que también se denominan **grados de libertad del numerador**. Se trata de los grados de libertad utilizados en la estimación intergrupal de la varianza, el numerador de la razón F. Los otros son los **grados de libertad dentro**, también denominados **grados de libertad del denominador**. Son los grados de libertad totales en el cálculo de la estimación intragrupal de la varianza, es decir, el denominador de la razón F.

[1] Es posible, por casualidad, que F sea mayor o menor a 1 en cualquier situación en particular. Tanto la intergrupal como la intragrupal son sólo estimaciones, y ambas pueden variar un poco aun cuando la hipótesis nula sea verdadera. Si F es considerablemente mayor a 1, rechazamos la hipótesis nula que establece que en realidad todas las poblaciones tienen la misma media. ¿Pero qué sucede si F es considerablemente menor a 1? Esto rara vez sucede. Cuando ocurre, podría indicar que existe menos variación entre los grupos de la que se esperaría por casualidad; por lo tanto, algo está restringiendo la variación entre los grupos. Una causa podría ser que, al organizar el experimento, se equiparen los sujetos entre los grupos en cuanto a determinadas variables (tales como edad o inteligencia) que resulten estar relacionadas con la variable bajo estudio. Una implicancia de esta posibilidad es que equiparar grupos de este modo, antes de realizar el estudio, podría realmente evitar un resultado significativo. Aun si existen diferencias reales entre las medias de la población, la influencia de estas diferencias en la estimación intergrupal puede ser compensada por el efecto de la equiparación. Este tema es tratado más adelante en el capítulo.

Los grados de libertad **entre** son la cantidad de grupos menos 1 (porque ese es el grado de libertad utilizado para calcular la estimación intergrupal de la varianza). Se expresan bajo la fórmula,

$$gl_{\text{Dentro}} = N_{\text{Grupos}} - 1 \tag{11-6}$$

Los grados de libertad **dentro** son la suma de los grados de libertad de todos los grupos (ya que todas sus estimaciones están incluidas en la combinación). Se expresan bajo la fórmula,

$$gl_{\text{Dentro}} = gl_1 + gl_2 + \ldots + gl_{\text{Último}} \tag{11-7}$$

En el ejemplo referido a los antecedentes delictivos, los grados de libertad entre son 2 (3 medias menos 1). Aplicando la fórmula,

$$gl_{\text{Entre}} = N_{\text{Grupos}} - 1 = 3 - 1 = 2.$$

Los grados de libertad dentro son 12, ya que cada grupo tiene 4 grados de libertad en los que se basa la estimación (5 registros menos 1) y hay 3 grupos en total que, sumados, dan como resultado 12 grados de libertad. Aplicando la fórmula,

$$gl_{\text{Entre}} = gl_1 + gl_2 + \ldots + gl_{\text{Último}} = (5-1) + (5-1) + (5-1) = 4 + 4 + 4 = 12.$$

Buscaríamos entonces el punto de corte en una distribución F "con 2 y 12" grados de libertad.

Como lo indica la tabla 11-6, con el nivel 0,05 se necesita una razón F de 3,89 para rechazar la hipótesis nula. (Con el nivel 0,01 se necesitaría una F de 6,93). La tabla F completa aparece con el nombre de tabla B-3 en el apéndice B.

Tabla 11-6.
Puntos de corte en la distribución F. (Información parcial).

Grados de libertad del denominador	Nivel de significación	Grados de libertad del numerador					
		1	*2*	*3*	*4*	*5*	*6*
10	0,01	10,05	7,56	6,55	6,00	5,64	5,39
	0,05	4,97	4,10	3,71	3,48	3,33	3,22
	0,10	3,29	2,93	2,73	2,61	2,52	2,46
11	0,01	9,65	7,21	6,22	5,67	5,32	5,07
	0,05	4,85	3,98	3,59	3,36	3,20	3,10
	0,10	3,23	2,86	2,66	2,54	2,45	2,39
12	0,01	9,33	6,93	5,95	5,41	5,07	4,82
	0,05	4,75	3,89	3,49	3,26	3,11	3,00
	0,10	3,18	2,81	2,61	2,48	2,40	2,33
13	0,01	9,07	6,70	5,74	5,21	4,86	4,62
	0,05	4,67	3,81	3,41	3,18	3,03	2,92
	0,10	3,14	2,76	2,56	2,43	2,35	2,28

Nota: La tabla completa es la tabla B-3 del apéndice B.

Analicemos cómo funcionan estos pasos en el experimento referido a los antecedentes delictivos.

1. **Replantear el problema en función de hipótesis de investigación e hipótesis nula de las poblaciones.** Existen tres poblaciones:

Población 1: jurados a quienes se informó que el acusado tenía antecedentes delictivos.
Población 2: jurados a quienes se informó que el acusado no tenía antecedentes.
Población 3: jurados a los que no se dio información acerca de los antecedentes del acusado.

La hipótesis nula establece que las tres poblaciones tienen la misma media ($\mu_1 = \mu_2 = \mu_3$). La hipótesis de investigación establece que las medias poblacionales son diferentes.

2. **Determinar las características de la distribución comparativa.** La distribución comparativa es una distribución F con 2 y 12 grados de libertad.

3. **Determinar el punto muestral de corte en la distribución comparativa, a partir del cual debería rechazarse la hipótesis nula.** Utilizando la tabla F para el nivel 0,05 de significación, la razón F necesaria es 3,89.

4. **Determinar el valor muestral en la distribución comparativa.** En el análisis de varianza, la distribución comparativa es una distribución F, y el valor muestral en esa distribución es, por lo tanto, la razón F. En el ejemplo, la razón F que calculamos era 4,07.

5. **Comparar los valores de los pasos 3 y 4 para decidir si se rechaza o no la hipótesis nula.** En el ejemplo que estamos analizando, la razón F calculada es más extrema que el punto de corte al nivel 0,05 de significación. Por lo tanto, el investigador rechazaría la hipótesis nula que establece que los tres grupos provienen de poblaciones con la misma media. El resultado sugiere que provienen de poblaciones con diferentes medias; que las personas expuestas a diferentes tipos de información (o a la falta de información) acerca de los antecedentes delictivos de un acusado, en una situación de este tipo calificarán de forma diferente al acusado en cuanto a su culpabilidad.[2]

Otro ejemplo

Mikulincer (1998) realizó una serie de estudios en Israel utilizando la misma medida de clasificación de estilo de vinculación que vimos anteriormente en este capítulo (*véase* tabla 11-1). Uno de sus estudios incluía a 30 alumnos universitarios (10 para cada estilo de vinculación), todos los cuales tenían relaciones amorosas serias. Como parte del estudio, cada noche cada alumno anotaba si durante el día su pareja había hecho algo que traicionara su confianza. Los participantes anotaban hechos tales como que su pareja llegaba muy tarde a un encuentro acordado o que "olvidaba" comentar al participante algún plan de importancia. Los resultados, junto con los cálculos del análisis de varianza, se indican en la tabla 11-7. A continuación, se detallan los pasos de la prueba de hipótesis.

[2] Varios estudios reales han investigado sobre el hecho de si conocer los antecedentes delictivos del acusado afecta la probabilidad de que sea condenado. En términos generales, la conclusión parece ser razonablemente coherente con la del estudio ficticio aquí descripto. Para una revisión de tales estudios, *véase* Dane y Wrightsman (1982).

Tabla 11-7.
Cantidad de hechos que traicionan la confianza de individuos "cometidos por sus parejas durante tres semanas" con tres estilos distintos de vinculación.

	Estilo de vinculación		
	Seguro	Evasivo	Ansioso-ambivalente
n	10	10	10
M	2,10	3,70	4,20
S	1,66	1,89	1,93
S^2	2,76	3,57	3,72

Distribución F:

$gl_{\text{Entre}} = N_{\text{Grupos}} - 1 = 3 - 1 = 2$

$gl_{\text{Dentro}} = gl_1 + gl_2 + \ldots + gl_{\text{Último}} = (10 - 1) + (10 - 1) + (10 - 1) = 9 + 9 + 9 = 27$

F necesaria con un nivel de significación de 0,05 tomado de la tabla F, $gl = 2, 27$: 3,36.

Estimación intergrupal de varianza poblacional:

Tabla en donde se encuentra la S^2 para las tres medias

	M	Desvío	Desvío cuadrático
Seguro	2,10	-1,23	1,51
Evasivo	3,70	0,37	0,14
Ansioso-ambivalente	4,20	0,87	0,76
	Σ: 10,00	$\Sigma\,(M - GM)^2$	2,41
	GM: 3,33		

$S^2_M = \Sigma\,(M - GM)^2/gl_{\text{Entre}} = 2{,}41/2 = 1{,}205$

S^2_{Entre} ó $MS_{\text{Entre}} = (S^2_M)\,(n) = (1{,}205)\,(10) = 12{,}05$

Estimación intragrupal de varianza poblacional:

$$S^2_{\text{Dentro}} \text{ ó } CM_{\text{Dentro}} = \frac{S^2_1 + S^2_2 + \ldots + S^2_{\text{último}}}{N_{\text{Grupos}}} = \frac{2{,}76 + 3{,}57 + 3{,}72}{3} = \frac{10{,}05}{3} = 3{,}35$$

Razón F: $F = S^2_{\text{Entre}}/S^2_{\text{Dentro}}$ ó $CM_{\text{Entre}}/CM_{\text{Dentro}} = 12{,}05/3{,}35 = 3{,}60$

Fuente: Mikulincer (1998).

1. **Replantear el problema en función de hipótesis de investigación e hipótesis nula de las poblaciones.** Existen tres poblaciones.

Población 1: alumnos con un estilo de vinculación seguro.
Población 2: alumnos con un estilo de vinculación evasivo.
Población 3: alumnos con un estilo de vinculación ansioso-ambivalente.

La hipótesis nula establece que estas tres poblaciones tienen la misma media ($\mu_1 = \mu_2 = \mu_3$). La hipótesis de investigación establece que sus medias son diferentes.

2. **Determinar las características de la distribución comparativa.** La distribución comparativa será una distribución F. Sus grados de libertad son calculados de la siguiente manera: la estimación de la varianza intergrupal se basa en tres grupos, dando 2 grados de libertad. La estimación intragrupal se basa en 9 grados de libertad (10 participantes) en cada uno de los tres grupos, dando un total de 27 grados de libertad.

3. **Determinar el punto muestral de corte en la distribución comparativa, a partir del cual debería rechazarse la hipótesis nula.** Utilizando la tabla B-3, en el apéndice B, buscamos

en la columna correspondiente a 2 grados de libertad en el numerador y nos detenemos en la fila correspondiente a 27 grados de libertad del denominador. Utilizaremos el nivel 0,05 de significación. El punto de corte F es de 3,36.

4. **Determinar el valor muestral en la distribución comparativa.** Este paso requiere la determinación de la razón F de la muestra. Calculamos el numerador, la estimación intergrupal de la varianza, en dos pasos. Primero estimamos la varianza de la distribución de medias de muestras con 10 observaciones. Lo hacemos utilizando las tres medias reales como si fueran números individuales. Tomamos sus desvíos cuadráticos con respecto a sus medias, lo que da un total de 2,41, y dividimos el resultado por los grados de libertad, que son 2. El cálculo da 1,205. El segundo paso es convertir esta estimación de varianza de una distribución de medias en una estimación de varianza de una población de observaciones individuales. Se logra multiplicando la varianza de la distribución de medias por el tamaño de muestra de cada media: multiplicamos 1,205 por 10. El resultado es una estimación de 12,05, que es el numerador de la razón F, es decir, la estimación de la varianza poblacional basada en la variación entre grupos.

El denominador de la razón F, es decir, la estimación intragrupal de varianza, es el promedio de las estimaciones de varianza poblacional calculado a partir de cada muestra. (En este punto debemos tener cuidado de no equivocarnos y utilizar estimaciones del desvío estándar de la población). En el estudio que estamos analizando, el promedio de 2,76, 3,57 y 3,72 es 3,35. Este es nuestro denominador de la razón F, es decir, la estimación de la varianza poblacional basada en la variación dentro de los grupos.

La razón F es la estimación intergrupal dividida por la estimación intragrupal, lo que da 3,60 (es decir, 12,05/3,35 = 3,60).

5. **Comparar los valores de los pasos 3 y 4 para decidir si se rechaza o no la hipótesis nula.** La razón F calculada es 3,60. Este resultado es más extremo que el punto de corte correspondiente a un nivel de significación 0,05, que es de 3,36. Por lo tanto, Mikulincer rechazó la hipótesis nula. Podía concluir que la cantidad de traiciones a su confianza, por parte de sus parejas, informada durante un período de 3 semanas por alumnos con los tres distintos estilos de vinculación, era diferente. Esta conclusión era coherente con la hipótesis de Mikulincer basada en teorías sobre la vinculación.

Resumen de los pasos a seguir para la prueba de hipótesis *utilizando el análisis de varianza*

La tabla 11-8 resume los pasos involucrados en un análisis de varianza del tipo que hemos estado estudiando en este capítulo.

SUPUESTOS DEL ANÁLISIS DE VARIANZA

Los supuestos del análisis de varianza son básicamente los mismos que los de la prueba t para medias independientes. Es decir, obtenemos resultados estrictamente precisos sólo cuando las poblaciones siguen una distribución normal y tienen la misma varianza. Además, al igual que con la prueba t, *en la práctica obtenemos resultados bastante aceptables aun cuando las poblaciones son* moderadamente distintas de lo normal y tienen diferencias moderadas en cuanto a las varianzas.

Como regla, podemos decir que si la estimación de varianza del grupo con la mayor estimación no es mayor a 4 ó 5 veces la varianza del grupo con la menor estimación, y los tamaños de muestra son iguales, las conclusiones deberían ser adecuadamente precisas.

Tabla 11-8.
Pasos a seguir para realizar un análisis de varianza (cuando los tamaños de muestra son iguales).

1. Replantear el problema en función de hipótesis de investigación e hipótesis nula de las poblaciones.

2. Determinar las características de la distribución comparativa.
   a) La distribución comparativa es una distribución F.
   b) Los grados de libertad del numerador son la cantidad de grupos menos 1: $gl_{\text{Entre}} = N_{\text{Grupos}} - 1$.
   c) Los grados de libertad del denominador son la suma de los grados de libertad de cada grupo (la cantidad en el grupo menos 1): $gl_{\text{Dentro}} = gl_1 + gl_2 + \ldots + gl_{\text{Último}}$.

3. Determinar el punto de corte en la distribución comparativa, a partir del cual debería rechazarse la hipótesis nula.
   a) Determinar el nivel de significación deseado.
   b) Buscar en una tabla F el punto de corte indicado, utilizando los grados de libertad calculados en el paso 2.

4. Determinar el valor muestral en la distribución comparativa. (Será una razón F).
   a) Calcular la estimación intergrupal de varianza poblacional (S^2_{Entre} ó CM_{Entre}).
      i) Calcular las medias de cada grupo.
      ii) Calcular una varianza estimada sobre la base de las medias de los grupos. $S^2_M = \Sigma(M - GM)^2/gl_{\text{Entre}}$
      iii) Convertir la estimación de la varianza de una distribución de medias en una estimación de la varianza de una población de individuos, multiplicándola por la cantidad de observaciones de cada grupo: S^2_{Entre} ó $CM_{\text{Entre}} = (S^2_M)(n)$.
   b) Calcular la estimación intragrupal de varianza poblacional (S^2_{Dentro} ó CM_{Dentro}).
      i) Calcular estimaciones de la varianza poblacional sobre la base de los valores observados de cada grupo: para cada grupo, $S^2 = \Sigma(X - M)^2/(n - 1) = SC/gl$.
      ii) Promediar las estimaciones de varianza: S^2_{Dentro} ó $CM_{\text{Dentro}} = (S^2_1 + S^2_2 + \ldots + S^2_{\text{Última}})/N_{\text{Grupos}}$
   c) Calcular la razón F: $F = S^2_{\text{Entre}}/S^2_{\text{Dentro}}$ ó $F = CM_{\text{Entre}}/CM_{\text{Dentro}}$

5. Comparar los valores de los pasos 3 y 4 para decidir si se rechaza o no la hipótesis nula.

Si el incumplimiento del supuesto de normalidad es grave, podemos utilizar procedimientos especiales que tienen en cuenta ese incumplimiento, ya sea transformando la distribución para hacerla más normal o bien no utilizando el supuesto de distribución normal. Presentaremos estos procedimientos en el capítulo 15. Si el incumplimiento del supuesto de igual varianza es serio, podemos utilizar procedimientos especiales que, de hecho, establecen un punto de corte de razón F más exigente para determinar si el resultado es significativo a determinado nivel (como por ejemplo el 5%). Estos procedimientos se describen en textos más avanzados.

TAMAÑO DEL EFECTO Y POTENCIA DEL ANÁLISIS DE VARIANZA

Tamaño del efecto

El tamaño del efecto en el caso del análisis de varianza es un poco más complejo que en una prueba t. Con la prueba t, sacamos la diferencia entre las dos medias y la dividimos por el desvío estándar. En el análisis de varianza, también podemos dividir por el desvío estándar; sin embargo, en ese caso tenemos más de dos medias, por lo cual no queda claro cuál es el equivalente a la diferencia entre medias, es decir, el numerador en el cálculo del tamaño del efecto. Cohen (1988) sugiere que en el análisis de varianza deberíamos considerar el tamaño del efecto como la variación entre medias. Específicamente, Cohen recomienda utilizar el desvío estándar de la distribución de medias. Así, define el **tamaño del efecto del análisis de varianza (f)** como el desvío estándar de la distribución de medias dividido por el desvío estándar de las observaciones individuales.[3] Para valores de población verdaderos conocidos (o predichos) la fórmula es la siguiente:

$$f = \frac{\sigma_M}{\sigma_{\text{Dentro}}} \qquad (11\text{-}8)$$

Para estimar el tamaño del efecto de un estudio ya realizado,

$$f = \frac{S_M}{S_{\text{Dentro}}} \qquad (11\text{-}9)$$

Las medidas de Cohen para el tamaño del efecto del análisis de varianza son: 0,10 para un efecto pequeño, 0,25 para un efecto mediano y 0,40 para un gran tamaño del efecto.

Analicemos nuestro experimento ficticio acerca de los antecedentes delictivos. En el estudio calculamos que S_M^2, la varianza estimada de la distribución de medias basada en las medias de las tres muestras, era igual a 4,34. S_M, la raíz cuadrada de S_M^2, es 2,08. Calculamos que S_{Dentro}^2, la estimación de la varianza de cada población de individuos, basada en las estimaciones de varianza utilizando los valores de cada grupo, era igual a 5,33. S_{Dentro}, la raíz cuadrada de S_{Dentro}^2, es 2,31. Aplicando la fórmula para el tamaño del efecto estimado a partir de un estudio completo,

$$f = \frac{S_M}{S_{\text{Dentro}}} = \frac{2{,}08}{2{,}31} = 0{,}90$$

Se trata de un tamaño del efecto muy grande (gracias a nuestros datos ficticios).

En el estudio referido a los estilos de vinculación realizado por Mikulincer (1998),

$$f = \frac{S_M}{S_{\text{Dentro}}} = \frac{1{,}098}{1{,}830} = 0{,}60$$

En este caso, también tenemos un gran tamaño del efecto. Tanto en el estudio acerca de los antecedentes delictivos como en el estudio realizado por Mikulincer, esperaríamos un tamaño del efecto importante sólo por saber que F fue significativa en un estudio con tamaños de muestra pequeños (*véase* capítulo 8).

También sucede que, con un poco de manipulación algebraica, podemos estimar el tamaño del efecto sólo conociendo F y la cantidad de observaciones en cada grupo. La fórmula es,

$$f = \frac{\sqrt{F}}{\sqrt{n}} \qquad (11\text{-}10)$$

[3] En el capítulo 12, después de que hayamos presentado el método del modelo estructural para el análisis de varianza, aprenderemos cómo calcular otro tipo de tamaño del efecto, la "proporción de varianza justificada". Este indicador del tamaño del efecto está relacionado con el mismo concepto en el análisis de regresión (capítulo 4), por eso tiene un significado más directo para muchos investigadores, y lo veremos con frecuencia. En el capítulo 12 trataremos la relación de f con este indicador del tamaño del efecto.

Por ejemplo, en el estudio acerca de los antecedentes delictivos habíamos calculado que F era 4,07, y había cinco personas en cada grupo. Utilizando la fórmula,

$$f = \frac{\sqrt{F}}{\sqrt{n}} = \frac{\sqrt{4{,}07}}{\sqrt{5}} = \frac{2{,}02}{2{,}24} = 0{,}90$$

Para el estudio acerca del estilo de vinculación realizado por Mikulincer (1998),

$$f = \frac{\sqrt{F}}{\sqrt{n}} = \frac{\sqrt{3{,}60}}{\sqrt{10}} = \frac{1{,}90}{3{,}16} = 0{,}60$$

(Los resultados son iguales a los que calculamos utilizando las estimaciones de varianza). La fórmula que acabamos de ver es muy útil cuando se evalúa el tamaño del efecto de un estudio en una publicación científica, en las que con frecuencia no se proporcionan detalles sobre las varianzas.

Potencia

La tabla 11-9 indica la potencia aproximada para un nivel de 0,05 de significación, con tamaños del efecto pequeños, medianos y grandes; tamaños de muestra de 10, 20, 30, 40, 50 y 100 por grupo, y para tres, cuatro o cinco grupos. Los anteriores son los valores más comunes de los distintos parámetros que influyen sobre la potencia.[4]

Analicemos un estudio planificado que compara cinco grupos de10 participantes cada uno, con expectativa de un gran tamaño del efecto (0,40), y que utiliza el nivel 0,05 de significación. Basándonos en la tabla 11-9, el estudio tendría una potencia de 0,56, lo que implica que aun si la hipótesis de investigación es en efecto verdadera y tiene un gran tamaño del efecto, existe sólo un poco más del 50 % de posibilidades (56%) de que el estudio resulte significativo.

Como observamos en capítulos anteriores, determinar la potencia es especialmente útil cuando se interpretan las implicancias prácticas de un resultado no significativo. Por ejemplo, supongamos que hemos leído un estudio que utiliza un análisis de varianza para cuatro grupos de 30 participantes cada uno, en el que el investigador informa un resultado no significativo al nivel 0,05 de significación. La tabla 11-9 indica una potencia de sólo 0,13 para un tamaño del efecto pequeño. Esto sugiere que aun si dicho efecto leve existe en la población, hubiera sido muy improbable que este estudio resultara significativo. Pero la tabla indica una potencia de 0,96 para un gran tamaño del efecto, lo que sugiere que si existiera un gran efecto en la población, casi seguramente se habría reflejado en el estudio.

Planificación del tamaño muestral

La tabla 11-10 nos indica la cantidad aproximada de participantes necesarios en cada grupo para tener un 80% de potencia al nivel 0,05 de significación, con tamaños estimados del efecto pequeños, medianos y grandes y en estudios con tres, cuatro y cinco grupos.[5] Por ejemplo, supongamos

[4] Cohen (1988, pp. 289-354) proporciona tablas más detalladas. Al utilizar estas tablas, se debe observar que el valor u en la parte superior de cada una de las tablas se refiere a gl_{Entre}, el cual en el caso de un análisis de varianza de un criterio es la cantidad de grupos menos 1, y no la cantidad de grupos, como sucede en la tabla 11-9.

[5] Cohen (1988, pp. 381-389) nos proporciona tablas más detalladas. Si se utilizan, se debe tener en cuenta la nota al pie número 4 de éste capítulo.

Tabla 11-9.
Potencia aproximada para estudios que utilizan el análisis de varianza probando la hipótesis a nivel 0,05 de significación.

	Tamaño del efecto		
Participantes por grupo (n)	Pequeño (0,10)	Mediano (0,25)	Grande (0,40)
Tres grupos ($gl_{Entre} = 2$)			
10	0,07	0,20	0,45
20	0,09	0,38	0,78
30	0,12	0,55	0,93
40	0,15	0,68	0,98
50	0,18	0,79	0,99
100	0,32	0,98	*
Cuatro grupos ($gl_{Entre} = 3$)			
10	0,07	0,21	0,51
20	0,10	0,43	0,85
30	0,13	0,61	0,96
40	0,16	0,76	0,99
50	0,19	0,85	*
100	0,36	0,99	*
Cinco grupos ($gl_{Entre} = 4$)			
10	0,07	0,23	0,56
20	0,10	0,47	0,90
30	0,13	0,67	0,98
40	0,17	0,81	*
50	0,21	0,90	*
100	0,40	*	*

*Casi 1.

que estamos planificando un estudio que involucra cuatro grupos, del cual esperamos un tamaño del efecto pequeño (y utilizáramos el nivel 0,05 de significación). Para obtener una potencia del 80% necesitaríamos 274 participantes en cada grupo, un total de 1.096. Sin embargo, supongamos que pudiéramos adaptar el plan de investigación de tal forma que fuera razonable predecir un gran tamaño del efecto (tal vez utilizando medidas más precisas y una manipulación experimental más poderosa). En ese caso, necesitaríamos sólo 18 participantes para cada uno de los cuatro grupos, un total de 72.

Tabla 11-10.
Cantidad aproximada de participantes necesarios en cada grupo (suponiendo que las muestras tengan el mismo tamaño) para lograr un 80% de potencia en un análisis de varianza de un criterio que prueba la hipótesis al nivel 0,05 de significación.

	Tamaño del efecto		
	Pequeño ($f = 0{,}10$)	Mediano ($f = 0{,}25$)	Grande ($f = 0{,}40$)
Tres grupos ($gl_{Entre} = 2$)	322	52	21
Cuatro grupos ($gl_{Entre} = 3$)	274	45	18
Cinco grupos ($gl_{Entre} = 4$)	240	39	16

CONTROVERSIAS Y LIMITACIONES: ASIGNACIÓN ALEATORIA VERSUS SELECCIÓN SISTEMÁTICA

Existe una controversia relacionada con el análisis de varianza que se refiere al diseño de experimentos. Comúnmente, la forma óptima de emprender un experimento es utilizando asignaciones totalmente aleatorias para las condiciones experimentales (*véase* apéndice A). Sin embargo, incluso con la asignación aleatoria continúa existiendo cierta variación aleatoria debido a las diferencias entre los participantes, lo cual agrega confusión al experimento. Por lo tanto, algunos investigadores modifican la asignación aleatoria estricta preparando sus estudios de modo de asegurarse que los participantes de cada grupo experimental sean, en líneas generales, semejantes en cuanto a una o más variables relevantes para el estudio. Analicemos un estudio en el que alumnos de cuarto grado serán asignados a uno de los tres diferentes programas experimentales de matemática. Es probable que los investigadores quieran asegurarse que el CI promedio y la capacidad promedio para la matemática sean iguales en cada uno de los grupos antes de comenzar el experimento.[6]

La controversia acerca de la utilización de equiparación de grupos, para minimizar las diferencias promedio en las variables relevantes, está relacionada con el efecto que ese procedimiento tiene sobre la potencia del análisis de varianza para probar el resultado del estudio. La selección sistemática reduce artificialmente la variación natural entre muestras (a tal punto que las variables sobre las cuales se realiza la equiparación están relacionadas con la variable estudiada). Si disminuimos la variación aleatoria entre las muestras, la variación general entre medias, es decir, el numerador de la razón F, en líneas generales sería menor. Por el contrario, el denominador de la razón F, la estimación intragrupal de la varianza, no es afectada por el hecho de que se realice una equiparación de grupos o una asignación aleatoria ordinaria. Si reducimos el numerador y mantenemos igual el denominador, la razón F sólo puede disminuir. Una razón F menor significa menor posibilidad de obtener un resultado significativo, aun si existe una verdadera diferencia de medias entre las poblaciones representadas por las condiciones experimentales; es decir, disminuye la potencia. (Lo que acabamos de describir se contradice con lo que nos diría nuestra intuición, ya que, a primera vista, reducir la "confusión" debería aumentar la potencia. El problema radica en que estamos reduciendo la confusión de manera despareja; por lo tanto, la confusión que normalmente contribuiría a la estimación intergrupal de la varianza se pierde, mientras que la

[6] Una forma de realizar esta selección sería comenzar con un grupo de todos los participantes disponibles. Primero seleccionaríamos al azar tantos como fueran necesarios para el primer grupo. Luego seleccionaríamos al azar alumnos adicionales para cada uno de los otros grupos, unos pocos por vez, adaptando las inclusiones coherentemente hasta que los tres grupos tuvieran los mismos promedios de CI y capacidad matemática. En este tipo de equiparación de grupos, la estructura resultante sigue siendo un verdadero experimento: el experimentador determina en qué grupo se incluye un participante utilizando procedimientos aleatorios; cualquier niño tiene las mismas posibilidades de pertenecer a cualquiera de los tres grupos. No se debe confundir esta clase de equiparación de grupos (la equiparación que vemos en esta sección) con otros dos tipos de equiparación. Uno de estos otros tipos de equiparación se realiza cuando la asignación aleatoria no es posible. Se intenta seleccionar personas de diferentes poblaciones preexistentes, de forma tal que las muestras sean lo más similares posibles. Un ejemplo sería un estudio en el que se comparen hombres y mujeres o personas de tres nacionalidades diferentes. En tal estudio, no podríamos asignar a las personas al azar a los grupos de los diferentes sexos o nacionalidades, pero podríamos intentar que los grupos que estudiamos sean similares en cuanto a la edad, preparación educativa, y así sucesivamente. Es un método mucho menos riguroso que la verdadera asignación aleatoria a los grupos.

Existe un segundo tipo de equiparación que no estamos tratando aquí, que es una especie de equiparación uno-a-uno. Por ejemplo, un investigador podría seleccionar series de tres estudiantes, en las que los tres estudiantes son muy similares y, luego, a partir de cada serie, los tres son asignados al azar a cualquiera de las tres condiciones experimentales. Este tipo de equiparación individual, que no es controvertida, es casi siempre ventajosa, pero rara vez práctica.

confusión que contribuye a la estimación intragrupal de varianza permanece igual). Por lo tanto, la recomendación tradicional en la mayoría de los libros de diseño experimental es que *no* se utilice este tipo de equiparación de grupos al programar los experimentos.

Sin embargo, Ross y Klein (1988) han cuestionado esta recomendación tradicional. Ellos reconocen que con la equiparación de grupos, el numerador de la razón F (y por lo tanto la razón F en su totalidad), en líneas generales se reduce. Pero también señalan que esto sucede en **líneas generales**, y que es bastante posible que, en determinadas situaciones que pueden especificarse, la razón F en realidad aumente por causa de ese procedimiento.

Ross y Klein realizaron una serie de estudios de Montecarlo (*véase* cuadro 10-1) para determinar el efecto real de la equiparación de grupos en distintas condiciones. El resultado de sus estudios fue que utilizar la equiparación de grupos, en comparación con la asignación aleatoria ordinaria, a) es conveniente si la hipótesis nula es verdadera, en cuanto a que se reduce la posibilidad de cometer un error Tipo I (rechazar equivocadamente la hipótesis nula); b) no es conveniente cuando la hipótesis de investigación es verdadera, pero las diferencias reales entre las medias grupales son pequeñas debido a que, en este caso, la potencia se reduce, y c) es conveniente cuando la hipótesis de investigación es verdadera y las diferencias reales entre las medias grupales son grandes porque, en este caso, la potencia aumenta. De todos modos, en todos los casos la mejor opción es utilizar la equiparación de grupos, pero analizando los resultados con un procedimiento estadístico más sofisticado denominado "análisis de covarianza" (brevemente descripto en el capítulo 17). El análisis de covarianza tiene en cuenta sistemáticamente los valores observados en cada participante en las variables en las que se realiza la equiparación. Lamentablemente, este procedimiento no puede ser utilizado en muchos casos, ya sea porque no pueden cumplirse los exigentes supuestos o porque la información necesaria no está disponible. Por lo tanto, cuando en un estudio es factible la equiparación de grupos, ésta parece recomendable en las situaciones a) y c) establecidas por Ross y Klein, aun si el procedimiento especial de análisis de covarianza no puede utilizarse, y el querido y viejo análisis de varianza estándar sí.

EL ANÁLISIS DE VARIANZA SEGÚN SE DESCRIBE EN LAS PUBLICACIONES CIENTÍFICAS

Un análisis de varianza del tipo que hemos analizado en este capítulo usualmente se describe en una publicación científica a través del F, los grados de libertad, y el nivel de significación, por ejemplo, "$F(3, 67) = 5{,}81$, $p < 0{,}01$." Las medias grupales generalmente se informan en una tabla, aunque si hay sólo unos pocos grupos y sólo una o unas pocas medidas, las medias pueden aparecer en el texto. Volviendo al ejemplo del experimento referido a los antecedentes delictivos, podríamos describir los resultados del análisis de varianza de la siguiente manera: "Las medias correspondientes al grupo de antecedentes delictivos, al grupo sin antecedentes delictivos, y al grupo sin información eran 7,0, 4,0 y 5,0 respectivamente, $F(2, 12) = 4{,}07$, $p < 0{,}05$."

El siguiente ejemplo fue tomado de una publicación. Grilo et al. (1997) son psicólogos clínicos interesados en la relación de la depresión y la utilización de substancias químicas con los trastornos de personalidad. Los trastornos de personalidad son características y comportamientos problemáticos y persistentes que exceden el rango usual de las diferencias individuales. Los investigadores realizaron entrevistas para evaluar trastornos de personalidad en adolescentes internados en hospitales psiquiátricos con alguno de estos tres diagnósticos: alto grado de depresión, abuso de substancias químicas y ambos, es decir, depresión y abuso de substancias químicas. Las cantidades medias de trastornos de personalidad fueron las siguientes: alto grado de depresión

$M = 1,0$; abuso de substancias $M = 0,7$; ambas condiciones $M = 1,9$. Los investigadores informaron: "Los tres grupos de estudio diferían en el promedio de trastornos de personalidad diagnosticados. $F(2,112) = 10,18$, $p < 0,0001$." En este estudio, como en la mayoría de los casos que involucran análisis de varianza, es común que los investigadores posteriormente realicen algún análisis sistemático para averiguar qué medias difieren significativamente de qué otras medias. Ese tema será tratado en el capítulo 12.

Resumen

El ANOVA prueba la hipótesis de que hay diferencias entre las medias de varias poblaciones. El procedimiento compara dos estimaciones de la varianza poblacional. Una, denominada "estimación intragrupal", que se determina por el promedio de las estimaciones de la varianza realizadas a partir de cada una de las muestras. La otra, denominada "estimación intergrupal", se basa en la variación entre las medias muestrales.

La razón F es igual a la estimación intergrupal dividida por la estimación intragrupal. La hipótesis nula establece que todas las muestras provienen de poblaciones con la misma media. Si la hipótesis nula es verdadera, la razón F debería ser aproximadamente 1, ya que las dos estimaciones de la varianza poblacional se basan en lo mismo, la variación dentro de cada población. Pero si la hipótesis de investigación es verdadera, y las muestras provienen de poblaciones con diferentes medias, la razón F debería ser mayor a 1, ya que la estimación intergrupal se ve, en ese caso, influenciada tanto por la variación dentro de las poblaciones como por la variación entre las poblaciones, mientras que la estimación intragrupal continúa afectada sólo por la variación dentro de cada una de las poblaciones.

Cuando las muestras tienen el mismo tamaño, la estimación intragrupal de la varianza poblacional es el promedio de las estimaciones de la varianza poblacional calculadas a partir de cada muestra. La estimación intergrupal de la varianza poblacional se realiza en dos pasos: primero, se estima la varianza de la distribución de medias sobre la base de las medias de las muestras reales (para realizar este cálculo se utiliza la fórmula usual de estimación de la varianza poblacional a partir de valores muestrales). En segundo lugar, se multiplica la estimación anterior por el tamaño de la muestra de cada grupo. A través de este segundo paso obtenemos la varianza de la distribución de valores individuales a partir de la varianza de la distribución de medias.

Los supuestos del análisis de varianza son los mismas que los de la prueba t: las poblaciones deben estar normalmente distribuidas y tener las mismas varianzas. Se ha descubierto que el análisis de varianza, al igual que la prueba t, otorga resultados razonablemente precisos aun cuando se violen moderadamente los supuestos.

El tamaño del efecto en el análisis de varianza puede calcularse como el desvío estándar de la distribución de medias dividido por el desvío estándar de la distribución de observaciones individuales. En el caso de un estudio ya realizado, también se puede calcular como la raíz cuadrada de F dividida por la raíz cuadrada de la cantidad de participantes en cada grupo. La potencia depende del tamaño del efecto, de la cantidad de personas que participan en el estudio, del nivel de significación y de la cantidad de grupos.

Asignar participantes sistemáticamente a los grupos experimentales, para asegurar promedios similares en cuanto a variables de fondo, generalmente reduce la potencia. Esto ocurre porque el procedimiento reduce la contribución de varianza aleatoria a la estimación intergrupal, pero no a la estimación intragrupal. Sin embargo, en ciertas condiciones el procedimiento puede aumentar la potencia.

Términos clave

- ANOVA.
- Grados de libertad intergrupales (gl_{Entre}).
- Estimación intergrupal de la varianza poblacional (S^2_{Entre} ó CM_{Entre}).
- Grados de libertad del denominador (gl_{Dentro}).
- Tamaño del efecto del análisis de varianza (f).
- Distribución F.
- Razón F.
- Tabla F.
- Gran media (GM).
- Grados de libertad del numerador (gl_{Entre}).
- Grados de libertad intragrupales (gl_{Dentro}).
- Estimación intragrupal de la varianza poblacional (S^2_{Dentro} ó CM_{Dentro}).

Ejercicios

Los ejercicios implican la realización de cálculos (con la ayuda de una calculadora). La mayoría de los problemas estadísticos reales se resuelven por computadora, pero aunque exista la posibilidad de utilizarla, es conveniente realizar estos ejercicios manualmente para incorporar el método de trabajo.

Para adquirir práctica en la utilización de una computadora, para resolver problemas estadísticos, se puede utilizar la sección de computación de cada capítulo, publicada en la *Guía de estudio y libro de tareas de computación para el alumno [Student's Study Guide and Computer Workbook]* que acompaña este libro.

Todos los datos de esta sección son ficticios (a menos que se especifique lo contrario)

Las respuestas a los ejercicios de la serie I se encuentran al final del libro.

SERIE I

1. Para cada uno de los siguientes estudios, decida si se puede rechazar la hipótesis nula que establece que los grupos provienen de poblaciones idénticas. Utilice el nivel 0,05. Además, calcule el tamaño del efecto y la potencia aproximada de cada uno. (Asegúrese de mostrar todos sus cálculos. Observe también que con respecto a los estudios b y c indicamos S, y no S^2).

(a)	Grupo 1	Grupo 2	Grupo 3
n	10	10	10
M	7,4	6,8	6,8
S^2	0,82	0,90	0,80

(b)	Grupo 1	Grupo 2	Grupo 3	Grupo 4
n	25	25	25	25
M	94	101	124	105
S	24	28	31	25

(c)	Grupo 1	Grupo 2	Grupo 3	Grupo 4	Grupo 5
n	25	25	25	25	25
M	94	101	124	105	106
S	24	28	31	25	27

2. Para cada uno de los siguientes estudios, decida si se puede rechazar la hipótesis nula que establece que los grupos provienen de poblaciones idénticas. Utilice el nivel 0,01. Además, calcule el tamaño del efecto y la potencia aproximada de cada uno. (Asegúrese de mostrar todos sus cálculos).

(a) Grupo 1	Grupo 2	Grupo 3
8	6	4
8	6	4
7	5	3
9	7	5

(b) Grupo 1	Grupo 2	Grupo 3
12	10	8
04	02	0
12	10	8
04	02	0

3. Se le pidió a un psicólogo de un hospital mental privado que determine si existía alguna diferencia clara en la duración del período de internación de pacientes con diferentes categorías de diagnóstico. Analizando a los últimos cuatro pacientes en cada una de las tres categorías más importantes, los resultados (en términos de semanas de internación) fueron los siguientes:

Categoría de diagnóstico

Trastornos afectivos	*Trastornos cognitivos*	*Trastornos relacionados con las drogas*
7	12	08
6	08	10
5	09	12
6	11	10

Utilizando un nivel 0,05, ¿existe alguna diferencia significativa en la duración del período de internación entre las distintas categorías de diagnóstico? a) Realice los cinco pasos de la prueba de hipótesis. b) Explique su respuesta a alguien que comprende todo lo relacionado con la realización de una prueba t para medias independientes pero que nunca ha escuchado hablar del análisis de varianza.

4. Un estudio comparaba la intensidad del amor no correspondido entre tres grupos: i) 50 individuos que estaban experimentando un amor no correspondido, que tenían una media de intensidad experimentada = 3,5, $S^2 = 5,2$; ii) 50 individuos que habían experimentado anteriormente un amor no correspondido y describían su experiencia retrospectivamente, $M = 3,2$, $S^2 = 5,8$ y iii) 50 individuos que nunca habían experimentado un amor no correspondido pero describían cómo pensaban que se sentirían si les ocurriera, $M = 3,8$, $S^2 = 4,8$. Determine la significación de la diferencia entre los grupos utilizando el nivel del 5%. a) Realice los cinco pasos de la prueba de hipótesis. b) Calcule el tamaño del efecto y la potencia. c) Explique su respuesta a alguien que nunca ha asistido a un curso de estadística.

5. Un investigador está preocupado porque considera que la necesidad cuidado de la salud mental entre los prisioneros difiere según el tipo de instalaciones de la prisión. El investigador selecciona al azar 40 prisioneros de cada uno de los tres tipos principales de prisión de un Estado determinado de los EE.UU. y realiza exámenes para determinar la necesidad de cuidado de la salud mental de los prisioneros. En la publicación que describe los resultados, el investigador informa las medias de cada grupo en cuanto a necesidad de cuidado de la salud mental, y luego agrega: "La necesidad de cuidado de la salud mental entre prisioneros de los tres tipos de sistemas penitenciarios parece ser claramente diferente, $F(2, 117) = 5,62$, $p < 0,01$." Explique el significado de lo anterior a una persona que nunca ha asistido a un curso sobre estadística.

6. ¿Qué clase de palabras son más largas, los sustantivos, los verbos o los adjetivos? Tome un diccionario, busque hojas al azar (utilizando los números aleatorios que aparecen debajo) y descienda por la columna hasta que encuentre un sustantivo. Anote su longitud (en cantidad de letras). Haga lo mismo con 10 sustantivos diferentes. Repita el proceso con 10 verbos y luego con 10 adjetivos. Después realice un análisis de varianza comparando los tres tipos de palabras. Además, suponiendo que existe una gran tamaño del efecto, ¿cuál es la potencia de este estudio (a un nivel de 0,05), y cuántas palabras de cada tipo serían necesarias para tener una potencia del 80%?.

651, 73, 950, 320, 564, 666, 736, 768, 661, 484, 990, 379, 323, 219, 715, 472, 176, 811, 167, 612, 102, 452, 849, 615, 228, 352, 851, 981, 821, 834, 719, 525, 907, 448, 4, 335, 671, 118, 403

SERIE II

1. Para cada uno de los siguientes estudios, decida si se puede rechazar la hipótesis nula que establece que los grupos provienen de poblaciones idénticas. Utilice el nivel de 0,05. Además, calcule el tamaño del efecto y la potencia aproximada de cada uno. (Asegúrese de mostrar todos sus cálculos).

(a) Grupo 1		Grupo 2	Grupo 3
n	5	5	5
M	10	12	14
S^2	4	6	5
(b) Grupo 1		**Grupo 2**	**Grupo 3**
n	10	10	10
M	10	12	14
S^2	4	6	5
(c) Grupo 1		**Grupo 2**	**Grupo 3**
n	5	5	5
M	10	14	18
S^2	4	6	5
(d) Grupo 1		**Grupo 2**	**Grupo 3**
n	5	5	5
M	10	12	14
S^2	2	3	2,5

2. Para cada uno de los siguientes estudios, decida si se puede rechazar la hipótesis nula que establece que los grupos provienen de poblaciones idénticas. Utilice el nivel 0,05. Además, calcule el tamaño del efecto para cada uno. (Asegúrese de mostrar todos sus cálculos).

(a) Grupo 1	Grupo 2	Grupo 3
1	1	8
2	2	7
1	1	8
2	2	7
(b) Grupo 1	**Grupo 2**	**Grupo 3**
1	4	8
2	5	7
1	4	8
2	5	7

3. Un psicólogo especializado en asuntos empresariales estaba interesado en averiguar si los individuos que trabajaban en diferentes sectores de la empresa tenían diferentes actitudes hacia la misma. Los resultados correspondientes a las tres personas entrevistadas del área de ingeniería fueron 10, 12 y 11; los resultados de los tres del área de comercialización 6, 6 y 8; los resultados de los tres miembros de contaduría, 7, 4 y 4; y los resultados de los tres de producción, 14, 16 y 13 (los números más altos indican actitudes más positivas). ¿Existía una diferencia de actitud significativa hacia la empresa entre empleados de diferentes sectores de la misma al nivel 0,05? a) Realice los cinco pasos de la prueba de hipótesis. b) Explique su respuesta a alguien que comprende todo lo relacionado con la realización de una prueba t para medias independientes pero que nunca ha escuchado hablar del análisis de varianza.

4. ¿Son diferentes los alumnos de distintas facultades en cuanto a su sociabilidad? Se seleccionaron al azar 25 alumnos de cada una de las tres facultades de una determinada ciudad, y se les pidió que informaran acerca de la cantidad de tiempo que dedicaban cada día a las relaciones sociales con otros alumnos. Los resultados para la facultad X fueron una media de 5 y una varianza poblacional estimada de 2; para la facultad Y, $M = 4$, $S^2 = 1{,}5$, y para la facultad Z, $M = 6$, $S^2 = 2{,}5$. ¿Cuál sería su conclusión? Utilice el nivel 0,05. a) Realice los cinco pasos de la prueba de hipótesis. b) Calcule el tamaño del efecto y la potencia. c) Explique su respuesta a alguien que nunca ha asistido a un curso de estadística.

5. Se realiza un experimento en el que 60 participantes completan una prueba de personalidad de cada uno, pero no acerca de la forma en que los participantes se ven a sí mismos. En realidad, se asignan 15 alumnos al azar para completar la prueba según el modo en que piensan que los ven sus madres (es decir, la forma en que creen que sus madres completarían la prueba para describir a los propios participantes); 15 para completarla según lo harían sus padres con respecto a ellos; 15 para completarla según lo harían sus mejores amigos con respecto a ellos, y 15 para completarla según lo harían sus profesores con respecto a ellos. Los resultados principales aparecen en la tabla 11-11. Explique estos resultados a una persona que nunca ha asistido a un curso de estadística.

6. Corte 100 papelitos de aproximadamente el mismo tamaño y escriba un uno en 16 papelitos, un dos en 34 papelitos, un tres en 34 papelitos, y un cuatro en 16 papelitos (está creando una distribución aproximadamente normal). Ponga los papelitos en un recipiente o sombrero, mézclelos, extraiga dos, anote los números que extrajo, y vuélvalos a introducir en el recipiente. Después extraiga otros dos,

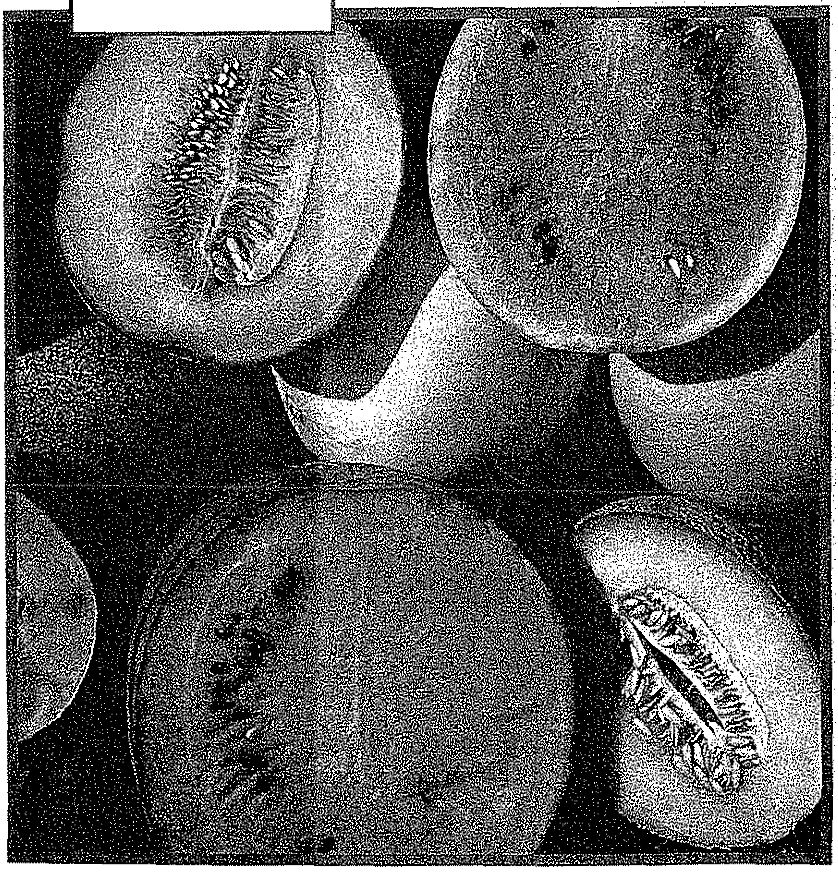
anote sus números, y vuélvalos a introducir; y finalmente otros dos, anote sus números y vuélvalos a introducir. (En un sentido estricto, debería realizar la muestra "con reemplazo"; eso significa volver a introducir cada uno de los papelitos después de anotar el número, y no los dos juntos, pero queremos ahorrarle un poco de tiempo, y en este caso la modificación del procedimiento no causará gran diferencia). Calcule un análisis de varianza para los tres grupos de dos valores cada uno, seleccionados al azar. Anote la razón *F* y repita todo el proceso de selección y análisis de varianza nuevamente. Realice todo el proceso al menos 20 veces y cree un polígono de frecuencias con sus resultados. Lo que está creando es una distribución *F* para 2 (3 grupos – 1) y 3 (4 – 1 en cada uno de los tres grupos) grados de libertad. ¿En qué punto comienza el 5% superior de sus valores *F*? Compare ese resultado con el punto de corte del 5% indicado en la tabla *F* del apéndice B para 2 y 3 grados de libertad.

Tabla 11-11.
Medias de las escalas principales de medición de la personalidad correspondientes a cada condición experimental (datos ficticios).

Escala	Madre	Padre	Amigo	Profesor	F (3, 56)
Conformidad	24	21	12	16	4,21**
Extroversión	14	13	15	13	2,05
Madurez	15	15	22	19	3,11*
Confianza en sí mismo	38	42	27	32	3,58*

*p < 0,05; **p < 0,01.

12 El modelo estructural en el análisis de varianza



Descripción del capítulo

En el capítulo 11 presentamos la lógica básica del análisis de varianza. A modo de revisión, podemos decir que el principio fundamental es que se realizan dos estimaciones de la varianza poblacional. Una, denominada estimación intergrupal de la varianza poblacional (S^2_{Entre} ó CM_{Entre}), se basa en la variación entre las medias de los grupos. La otra, denominada estimación intragrupal de la varianza poblacional (S^2_{Dentro} ó CM_{Dentro}), se basa en la variación de los registros dentro de cada uno de los grupos. Si la hipótesis nula es verdadera, las dos estimaciones de la varianza poblacional deberían ser aproximadamente iguales y, por ende, la razón entre la estimación intergrupal y la estimación intragrupal, es decir, la razón F, debería ser aproximadamente 1. En cambio, cuando la hipótesis nula es falsa, la estimación intergrupal estará influida por la diferencia entre las medias poblacionales. Por lo tanto, la estimación intergrupal será mayor que la intragrupal, y la razón F será mayor a 1. En la prueba de hipótesis comparamos la razón F calculada con un punto de corte (obtenido de la tabla F). El punto de corte es el extremo inferior de un intervalo de valores mayores que 1, el cual se extiende sin límite. La probabilidad de obtener una razón F en ese intervalo es del 5% (o del 1%) si la hipótesis nula es verdadera.

Partiendo de esta base, en el capítulo 12 exploramos una forma alternativa, pero matemáticamente equivalente, de interpretar el análisis de varianza. Esa alternativa se denomina **modelo estructural**. Si bien también se aplica la lógica central aprendida en el capítulo 11, el modelo estructural proporciona una forma diferente y más flexible de calcular las dos estimaciones de varianza poblacional. Este nuevo método facilita el manejo de aquella situación en la que la cantidad de individuos de cada grupo no es la misma, situación especial que analizamos en este capítulo. Además, al comprender el modelo estructural podremos entender con mayor profundidad la lógica implícita del análisis de varianza. Finalmente, la comprensión del método del modelo estructural ayudará a entender la forma en que las computadoras presentan los resultados del análisis de varianza.

PRINCIPIOS DEL MODELO ESTRUCTURAL

Partición de la desviación

La idea central del modelo estructural requiere pensar en términos de desviación. En primer lugar, existe la desviación de una observación con respecto a la gran media. La gran media es la media de **todas** las observaciones, independientemente del grupo en el que se encuentran. En el ejemplo del estudio acerca de los antecedentes delictivos, analizado en el capítulo 11, la gran media de los 15 valores observados era 85/15 = 5,67. En el ejemplo del estudio referido a estilos de vinculación, estudiado en el mismo capítulo, la gran media de las 30 observaciones con respecto a la traición de la confianza era 3,33.

Después debemos pensar que la desviación con respecto a la gran media tiene dos partes: a) la desviación de la observación con respecto a la media de su grupo y b) la desviación de la media de su grupo con respecto a la gran media. Analicemos a un participante en el estudio acerca de los antecedentes delictivos que calificó la culpabilidad del acusado con 10. La gran media de las calificaciones de culpabilidad de todos los participantes era 5,67. La calificación de la persona en cuestión presenta una desviación total de 4,33 (10 – 5,67 = 4,33). La media, únicamente del grupo de antecedentes delictivos, era 8. Por lo tanto, *la desviación de la calificación de esta persona* con respecto a la media de su grupo es 2 (es decir, 10 – 8 = 2), y la desviación de la media grupal con respecto a la gran media es 2,33 (es decir, 8 – 5,67 = 2,33). Es importante observar que esas dos desviaciones (2 y 2,33) suman la desviación total de 4,33. La figura 12-1 grafica lo anterior. Es conveniente estudiar este concepto hasta comprenderlo bien.

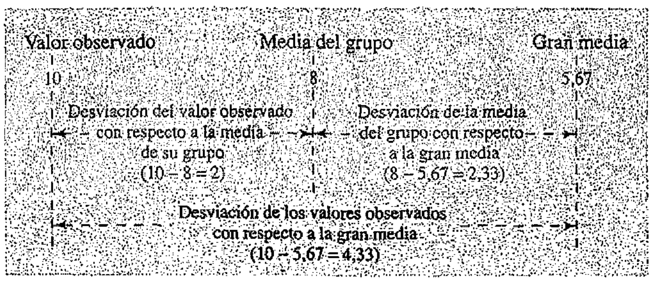


Figura 12-1. Ejemplo tomado de un estudio ficticio acerca de antecedentes delictivos, en el que se representa la desviación de la observación de un individuo con respecto a la gran media como la suma de la desviación de la observación individual con respecto a la media de su grupo, más la desviación de la media de su grupo con respecto a la gran media.

Suma de las desviaciones cuadráticas

El siguiente paso en la utilización de estas diferentes desviaciones es elevar a cada una al cuadrado y sumar cada tipo de desviaciones cuadráticas de todos los participantes. El resultado es la **suma de desviaciones cuadráticas** de cada tipo de desviación. Sucede que la suma de desviaciones cuadráticas de cada observación con respecto a la gran media es igual a a) la suma de las desviaciones cuadráticas de cada observación con respecto a la media de su grupo más b) la suma de las desviaciones cuadráticas de la media del grupo de cada observación con respecto a la gran media.

El principio que acabamos de explicar se puede expresar con una fórmula:

$$\begin{aligned}&\Sigma(X-GM)^2=\Sigma(X-M)^2+\Sigma(M-GM)^2\\ &\quad \text{ó } SC_{\text{Total}}=SC_{\text{Dentro}}+SC_{\text{Entre}}\end{aligned} \qquad (12\text{-}1)$$

En esta fórmula, $\Sigma\ (X - GM)^2$ ó SC_{Total} es la **suma de desviaciones cuadráticas** de cada observación con respecto a la gran media, sin tener en cuenta el grupo en el que se encuentra la observación. $\Sigma\ (X - M)^2$ ó SC_{Dentro} es la suma de la desviación cuadrática de cada observación con respecto a la media de su grupo, sumado al de todos los participantes. $\Sigma\ (M - GM)^2$ ó SC_{Entre} es la suma de la desviación cuadrática de la media del grupo al que pertenece cada observación con respecto a la gran media (nuevamente, en la suma interviene la de todos los participantes).

Esta regla se aplica sólo a las **sumas** de las desviaciones cuadráticas. Si tomamos las observaciones individualmente, la suma de las desviaciones siempre coincide, pero no la de las desviaciones cuadráticas.

Estimaciones de la varianza poblacional a partir de la suma de desviaciones cuadráticas

Ahora estamos listos para utilizar las sumas de las desviaciones cuadráticas para calcular las dos estimaciones de la varianza poblacional necesarias. Para hacerlo, dividimos cada suma de desviaciones cuadráticas por los grados de libertad correspondientes. La estimación intergrupal de la varianza poblacional (S^2_{Entre} ó CM_{Entre}) es la suma de las desviaciones cuadráticas de la media grupal de cada observación con respecto a la gran media (SC_{Entre}), dividida por los grados de libertad en los que se basa (gl_{Entre}, la cantidad de grupos menos 1). Lo anterior se expresa bajo la siguiente fórmula,

$$S^2_{\text{Entre}}=\frac{\Sigma(M-GM)^2}{gl_{\text{Entre}}}\ \text{ó}\ CM_{\text{Entre}}=\frac{SC_{\text{Entre}}}{gl_{\text{Entre}}} \qquad (12\text{-}2)$$

La estimación intragrupal de la varianza poblacional (S^2_{Dentro} ó CM_{Dentro}) es la suma de las desviaciones cuadráticas de cada observación con respecto a la media de su grupo (SC_{Dentro}), dividida por los grados totales de libertad en los que se basa (gl_{Dentro}; la suma de los grados de libertad de todos los grupos, es decir, la cantidad de observaciones del primer grupo menos 1, más la cantidad de observaciones del segundo grupo menos 1, etc.) Lo anterior se expresa bajo la fórmula,

$$S^2_{\text{Dentro}}=\frac{\Sigma(X-M)^2}{gl_{\text{Dentro}}}\ \text{ó}\ CM_{\text{Dentro}}=\frac{SC_{\text{Dentro}}}{gl_{\text{Dentro}}} \qquad (12\text{-}3)$$

Es importante mencionar que hemos ignorado la suma de desviaciones cuadráticas de cada observación con respecto a la gran media (SC_{Total}). Esta suma de cuadrados es útil principalmente para controlar nuestra aritmética. Recordemos que $SC_{\text{Total}} = SC_{\text{Dentro}} + SC_{\text{Entre}}$.

La figura 12-2 nuevamente representa la partición de la desviación en dos partes, pero esta vez acentúa la estimación de la varianza poblacional con la que se relaciona cada desviación.

Relación del enfoque del modelo estructural con el método del capítulo 11

Los métodos que acabamos de describir para el cálculo de las estimaciones intragrupal e intergrupal de la varianza, utilizando el modelo estructural, dan exactamente el mismo resultado que los métodos que aprendimos en el capítulo 11. (Si al alumno le divierte realizar manipulaciones algebraicas, podría intentar llegar a las fórmulas anteriores a partir de las que acabamos de aprender).

De todos modos, los procedimientos que realizamos para calcular esas estimaciones son bastante diferentes. En el caso del enfoque del modelo estructural que vemos en este capítulo, cuando desarrollamos el método de estimación intragrupal de la varianza, en realidad nunca calculamos la estimación de varianza de cada grupo para luego promediarlas. Del mismo modo, en el caso de la estimación intergrupal, con el método del modelo estructural, nunca multiplicamos ningún número por la cantidad de observaciones de cada muestra. Sin embargo, lo importante es que, con ambos métodos, obtenemos las mismas estimaciones intragrupal e intergrupal de varianza. Por lo tanto, de cualquier modo, los componentes utilizados para calcular la razón F son los mismos. Y de todas maneras, el resultado es el mismo.

La lógica implícita en el análisis de varianza con el modelo estructural es también esencialmente la misma que la que aprendimos en el capítulo 11, pero con un pequeño cambio. Lo que permanece igual es el hecho de que si la hipótesis nula es verdadera, las dos estimaciones de la varianza poblacional deberían ser prácticamente iguales, y si la hipótesis nula es falsa, la estimación intergrupal debería ser mayor (porque las diferencias entre las medias poblacionales contribuyen a ello) que la estimación intragrupal. El cambio radica en el énfasis. El método que aprendimos en el capítulo 11 hace hincapié en los grupos íntegros, comparando una varianza basada en las diferencias entre medias grupales con una varianza basada en el promedio de las varianzas grupales. El modelo estructural pone el acento en las observaciones individuales. Compara una varianza basada en las desviaciones de las medias de los grupos a los que pertene-

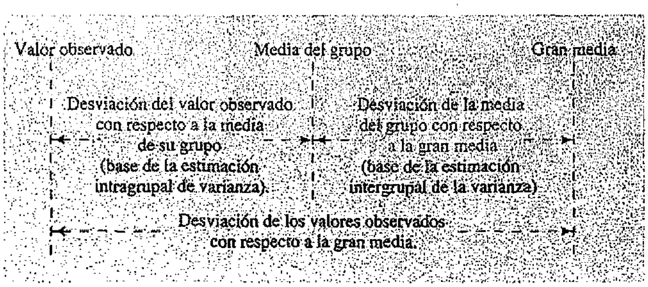


Figura 12-2. La desviación de las observaciones con respecto a la media de su grupo es la base para la estimación intragrupal de la varianza poblacional. La desviación de la media del grupo con respecto a la gran media es la base de la estimación intergrupal de la varianza poblacional.

cen los valores observados individuales con respecto a la gran media, con una varianza basada en los desvíos de los valores observados individuales con respecto a la media de su grupo. El método del capítulo 11 se concentra directamente en los aspectos que contribuyen a la estimación general de la varianza poblacional; el modelo estructural se concentra directamente en los aspectos que contribuyen a las partes en que se descomponen las desviaciones de las observaciones con respecto a la gran media.

Las diferencias lógicas mencionadas anteriormente son bastante sutiles y, finalmente, se reducen a lo mismo. Entonces, si tanto los cálculos como la lógica se refieren a lo mismo, ¿por qué debemos aprender dos formas diferentes de razonar ese tema? Hemos analizado el método del capítulo 11 principalmente porque es más intuitivo. Es especialmente útil para ayudar a comprender de qué se tratan las estimaciones de la varianza poblacional, y por qué deberían ser iguales cuando la hipótesis nula es verdadera y diferentes cuando no lo es. Además, con el método del capítulo 11 podemos calcular un análisis de varianza en forma directa a partir de medias y varianzas de grupos, sin necesidad de trabajar directamente con las observaciones.

Sin embargo, como dijimos al comienzo del capítulo, es importante presentar el modelo estructural porque a) ha sido el más utilizado (en parte porque es más cercano a las fórmulas de cálculo que durante tanto tiempo dominaron el razonamiento de todos), b) es más flexible, y por lo tanto más fácil de utilizar cuando se trabaja con grupos de tamaños desiguales y con el análisis factorial de varianza (presentado en el capítulo 13) y c) está relacionado con un método matemático fundamental que queríamos estar seguros de exponer a aquellos alumnos que podrían llegar a asistir a cursos más avanzados de estadística.

Cuadro 12-1. El análisis de varianza como forma de pensar acerca del mundo.

El análisis de varianza es una maravillosa idea básica que vale la pena analizar un poco más, por dos motivos: primero, porque en la medida que el alumno lea o realice investigaciones irá progresivamente adoptando esa manera de razonar y, segundo, porque de hecho es la forma en la que ya piensa.

Al realizar cualquier investigación (o al intentar decidir la calidad de un estudio que estamos leyendo) lo que nos interesa saber es si determinada variable realmente origina alguna diferencia. Organizamos dos (o más) grupos para poder mostrar que cualquier diferencia en los resultados existe puramente porque un grupo recibió la influencia de la variable y el otro no. (*Véase* apéndice A).

Por ejemplo, para observar el efecto de iniciar una conversación amistosa, hacemos que un grupo converse con un extraño cada día durante una semana y que otro grupo no haga nada en especial. Después buscamos una diferencia en cuanto a la cantidad de nuevas amistades. Para estar seguros de si se produjo alguna diferencia, uniformamos en los dos grupos la influencia de cualquier otro factor que pudiera causar esa misma diferencia: nadie debe asociarse a un club o ir a fiestas durante esa semana. También intentamos controlar cualquier otro efecto "aleatorio" que pudiera provocar suficiente variación como para facilitar que un grupo presentara diferencias por razones accidentales, y no utilizamos personas muy diferentes a lo normal en cuanto a su atractivo físico o que no hablen el idioma con fluidez. (El procedimiento anteriormente descripto no es igual que la equiparación de grupos, discutida en el capítulo 11, según la cual los grupos son iguales con respecto a alguna variable).

Este tipo de razonamiento estándar acerca del diseño de la investigación es paralelo a la lógica del análisis de varianza, una técnica puramente estadística. Como lo expresa uno de los libros de texto clásicos sobre diseño de investigación (Derlinger, 1973): "La principal función *técnica* del diseño de la investigación es controlar la varianza" (p. 306). Es decir, los investigadores buscan maximizar la varianza de la variable expresada por la hipótesis de investigación (el numerador, o varianza intergrupal), y controlar las variables accidentales que no están bajo estudio (las que contribuyen al denominador, o varianza intragrupal). Por lo tanto, el análisis de la varianza es muy similar a la forma de pensar cuando planificamos un experimento.

También dijimos, sin embargo, que el análisis de varianza es similar a la forma en que siempre hemos pensado. Kelley (1971) sugirió que, en el fondo, todos somos científicos, puesto que todos formulamos hipótesis y las sometemos a prueba; y el método que utilizamos para distinguir y tomar decisiones acerca de la causalidad aplica el razonamiento del análisis de varianza. Supongamos que estamos de visita en un país, y mientras viajamos observamos a una mujer rubia que arroja cartas dentro de una caja azul brillante con forma abovedada. Si vemos otras mujeres rubias arrojando cartas dentro de cajas verdes, cajones rojos, cilindros violetas y latas amarillas, construimos la información de forma tal que las variables importantes sean el color de pelo y el sexo de la persona (convertimos a la persona en el numerador, la varianza intergrupal). El recipiente puede variar y no es importante (lo convertimos en el denominador, la varianza intragrupal). Pero si vemos muchos otros tipos de personas poniendo cartas siempre en cajas abovedadas color azul brillante, sabemos que las cajas son la diferencia intergrupal que importa, y las personas sólo elementos aleatorios (en Canadá, serían cajas rojas).

Del mismo modo, como ingenuos psicólogos podemos interesarnos por el concepto denominado honestidad. ¿Es una característica de ciertas personas o "todos tienen su precio"? Sin duda, cada uno habrá observado a distintas personas y situaciones durante su vida, y tendrá su propia teoría al respecto. Esa teoría refleja qué aspecto consideramos mayor, es decir, el numerador, la característica de honestidad, o el denominador, el efecto de situaciones (tales como el monto de un soborno, la posibilidad de que los demás se enteren si uno fue honesto, etcétera).

Si el alumno está familiarizado con el trabajo del psicólogo Jean Piaget, especialista en el campo del desarrollo, reconocerá que el tipo de razonamiento del análisis de varianza es parte de lo que él llamaba "operaciones formales", el estilo de pensamiento abstracto normalmente adquirido alrededor de los 14 años. Por lo tanto, no deberíamos tener inconvenientes en comprender el análisis de varianza, ¡inocentemente, lo hemos estado utilizando durante años!

UTILIZACIÓN DEL MODELO ESTRUCTURAL PARA REALIZAR UN ANÁLISIS DE VARIANZA

La tabla 12-1 indica los cálculos completos utilizando el modelo estructural para realizar un análisis de la varianza del estudio acerca de los antecedentes delictivos. La tabla muestra los tres tipos de desviaciones y desviaciones cuadráticas de cada observación. Por ejemplo, en el caso de la primera persona, la desviación con respecto a la gran media es 4,33 (el 10 menos la gran media de 5,67), y esa desviación elevada al cuadrado es 18,74. La desviación de la observación con respecto a la media de su grupo es 2, y esa desviación elevada al cuadrado es 4. Finalmente, la desviación de la media del grupo de la observación con respecto a la gran media es 2,33, y esa desviación elevada al cuadrado es 5,43. Cabe notar que la desviación de la media del grupo de cada observación con respecto a la gran media (en este caso 2,33) es igual para todas las observaciones de un mismo grupo. Al final de cada columna, también hemos sumado las desviaciones cuadráticas de cada tipo.

La parte inferior de la tabla 12-1 indica los cálculos del análisis de varianza. Primero, calculamos las tres sumas de desviaciones cuadráticas (SC_{Total}, SC_{Dentro} y SC_{Entre}). El siguiente paso es controlar la exactitud del cálculo; lo hacemos siguiendo el principio que establece que la suma de las desviaciones cuadráticas de cada observación con respecto a la gran media es igual a la suma de los otros dos tipos de desviaciones cuadráticas.

Tabla 12-1.
Análisis de varianza del estudio acerca de los antecedentes delictivos (datos ficticios) utilizando el método del modelo estructural (comparar con tablas 11-3 y 11-4).

Grupo con antecedentes delictivos

X	$X - GM$		$X - M$		$M - GM$	
	Desviación	*Desviación cuadrática*	*Desviación*	*Desviación cuadrática*	*Desviación*	*Desviación cuadrática*
10	4,33	18,74	2	4	2,33	5,43
7	1,33	1,77	−1	1	2,33	5,43
5	−0,67	0,45	−3	9	2,33	5,43
10	4,33	18,74	2	4	2,33	5,43
8	2,33	5,43	0	0	2,33	5,43
40		45,13		18		27,14

$M = 40/5 = 8$

Grupo sin antecedentes delictivos

X	$X - GM$		$X - M$		$M - GM$	
	Desviación	*Desviación cuadrática*	*Desviación*	*Desviación cuadrática*	*Desviación*	*Desviación cuadrática*
5	−0,67	0,45	1	1	−1,67	2,79
1	−4,67	21,81	−3	9	−1,67	2,79
3	−2,67	7,13	−1	1	−1,67	2,79
7	1,33	1,77	3	9	−1,67	2,79
4	−1,67	2,79	0	0	−1,67	2,79
20		33,95		20		13,95

$M = 20/5 = 4$

Tabla 12-1 (*cont.*).

Grupo sin información sobre antecedentes						
X	$X - GM$		$X - M$		$M - GM$	
	Desviación	Desviación cuadrática	Desviación	Desviación cuadrática	Desviación	Desviación cuadrática
4	−1,67	2,79	−1	1	−0,67	0,45
6	0,33	0,11	1	1	−0,67	0,45
9	3,33	11,09	4	16	−0,67	0,45
3	−2,67	7,13	−2	4	−0,67	0,45
3	−2,67	7,13	−2	4	−0,67	0,45
25		28,25		26		2,25

$M = 25/5 = 5$

Sumas de desviaciones cuadráticas:

$\Sigma(X - GM)^2$ ó $SC_{\text{Total}} = 45{,}13 + 33{,}95 + 28{,}25 = 107{,}33$

$\Sigma(X - M)^2$ ó $SC_{\text{Dentro}} = 18 + 20 + 26 = 64$

$\Sigma(M - GM)^2$ ó $SC_{\text{Entre}} = 27{,}14 + 13{,}95 + 2{,}25 = 43{,}34$

Control ($SC_{\text{Total}} = SC_{\text{Dentro}} + SC_{\text{Entre}}$):

$SC_{\text{Total}} = 107{,}33$; $SC_{\text{Dentro}} + SC_{\text{Entre}} = 64 + 43{,}34 = 107{,}34$

(leve diferencia debido a error de redondeo)

Grados de libertad:

$gl_{\text{Total}} = N - 1 = 15 - 1 = 14$

$gl_{\text{Dentro}} = gl_1 + gl_2 + \ldots + gl_{\text{Último}} = (5 - 1) + (5 - 1) + (5 - 1) = 4 + 4 + 4 = 12$

$gl_{\text{Entre}} = N_{\text{Grupos}} - 1 = 3 - 1 = 2$

Control ($gl_{\text{Total}} = gl_{\text{Dentro}} + gl_{\text{Entre}}$): $14 = 12 + 2$

Estimaciones de varianza poblacional:

S^2_{Dentro} ó $CM_{\text{Dentro}} = SC_{\text{Dentro}}/gl_{\text{Dentro}} = 64/12 = 5{,}33$

S^2_{Entre} ó $CM_{\text{Entre}} = SC_{\text{Entre}}/gl_{\text{Entre}} = 43{,}34/2 = 21{,}67$

Razón F: $F = S^2_{\text{Entre}}/S^2_{\text{Dentro}}$ ó $CM_{\text{Entre}}/CM_{\text{Dentro}} = 21{,}67/5{,}33 = 4{,}07$

Los grados de libertad, el siguiente paso que aparece en la tabla, se calculan de la misma forma que en el capítulo 11. Más abajo, la tabla indica los cálculos de las dos estimaciones cruciales de varianza poblacional. Las calculamos dividiendo cada suma de desviaciones cuadráticas por los grados de libertad correspondientes. Finalmente, la tabla muestra el cálculo de la razón F, realizado de forma usual, dividiendo la estimación intergrupal de la varianza por la estimación intragrupal de varianza. Todos esos números, grados de libertad, estimaciones de varianza y F son iguales (con diferencias de redondeo) a las cifras calculadas en el capítulo 11.

TABLAS DE ANÁLISIS DE VARIANZA

Una **tabla de análisis de varianza** presenta los resultados de un análisis de varianza basándose en el enfoque del modelo estructural. Estas tablas son producidas automáticamente por la mayoría de los programas de análisis de varianza para computadoras. Una tabla estándar de análisis de varianza tiene cinco columnas. La primera columna está encabezada, por lo general, con el título "Fuente", y contiene el tipo de estimación de varianza o desvío involucrado ("intergrupal", "intragrupal" y "total"). La siguiente columna se titula habitualmente "*SC*", y contiene los diferentes tipos de sumas de desviaciones cuadráticas. La tercera columna es "*gl*", y contiene los distintos tipos de grados de libertad. La cuarta columna es "*CM*", que se refiere a los cuadrados medios. Es decir, *CM* es *SC* dividido por *gl*, la estimación de la varianza. *CM* es, como siempre, lo mismo que S^2. Sin embargo, en una tabla de análisis de varianza, la varianza casi siempre es representada como *CM*. La última columna es "*F*", la razón *F*. Cada fila de la tabla se refiere a una de las estimaciones de varianza. La primera fila corresponde a la estimación intergrupal de varianza. Generalmente aparece debajo de la columna "Fuente" como "intergrupal" o "grupal", aunque algunas veces se la denomina "modelo" o "tratamiento". La segunda fila corresponde a la estimación intragrupal de varianza, aunque algunas veces se la denomina "error". La última fila es la suma de los cuadrados sobre la base de la desviación total de cada observación con respecto a la gran media.

La tabla 12-2 es una tabla de análisis de varianza completa con los datos tomados del ejemplo referido al estudio de los antecedentes delictivos. Los diseños realizados por computadora algunas veces utilizan un orden diferente para las columnas y omiten *SC* ó *CM*, pero nunca los dos.

Tabla 12-2.
Tabla de análisis de varianza correspondiente al estudio acerca de antecedentes delictivos. (datos ficticios).

Fuente	*SC*	*gl*	*CM*	*F*
Intergrupal	43,34	2	21,67	4,07
Intragrupal	64	12	5,33	
Total	107,33	14		

ANÁLISIS DE VARIANZA CON GRUPOS DE TAMAÑOS DESIGUALES

Ya sea que los grupos tengan la misma cantidad de valores observados o no, la lógica básica del análisis de varianza es la misma. En ambos casos, es una comparación de estimaciones de la varianza poblacional sobre la base de la variación intergrupal versus la intragrupal. Sin embargo, los procedimientos para el cálculo de estimaciones intragrupales e intergrupales de la varianza, que aprendimos en el capítulo 11, son bastante difíciles de utilizar con grupos de tamaños desiguales.

Los procedimientos requieren adaptaciones complejas para dar la ponderación adecuada a la información obtenida de los grupos de tamaños desiguales. Por otro lado, el método del modelo estructural funciona de forma tal que, automáticamente, realiza las adaptaciones necesarias para grupos de tamaños desiguales.[1]

[1] Una ventaja del método que aprendimos en el capítulo 11, además de su utilidad para clarificar la lógica implícita, es que permite calcular un análisis de varianza utilizando sólo las medias y las varianzas poblacionales estimadas. Esto puede resultar útil cuando la información ordinaria no está disponible; por ejemplo, al calcular un análisis de varianza basándonos en medias y desvíos estándar informados en una publicación científica. Por lo tanto, si los tamaños de las muestras no fueran iguales, la siguiente es la forma de determinar las estimaciones de varianza poblacional utilizando el método del capítulo 11.

El cálculo de S^2_{Dentro} con distintos tamaños de muestra es una extensión directa del método que aprendimos en el capítulo 10 para $S^2_{\text{Combinada}}$, la estimación combinada de la varianza poblacional. Es decir, S^2_{Dentro} es la suma de la S^2 ponderada de cada grupo, siendo la ponderación los *gl* del grupo en cuestión (su *N* menos 1) divididos por los *gl* totales de todos los grupos. Es decir,

Ejemplo

Analicemos un ejemplo ficticio. Un investigador de un centro de tratamiento del alcoholismo realiza un estudio acerca de la satisfacción del paciente con tres métodos diferentes de tratamiento utilizados en el centro. Los llamaremos tratamiento A, tratamiento B y tratamiento C. El investigador asigna al azar a cada uno de los 10 pacientes disponibles para que reciban uno de estos tratamientos: a 4 pacientes les toca el tratamiento A, a 3 pacientes el tratamiento B y a 3 pacientes el tratamiento C. Dos semanas más tarde, el investigador mide la satisfacción de los pacientes con respecto a los tres tratamientos en una escala del 1 (bajo nivel de satisfacción) al 20 (alto nivel de satisfacción). La tabla 12-3 muestra los resultados, los cálculos y la tabla del análisis de varianza. La figura 12-3 representa gráficamente las distintas distribuciones involucradas. Seguiremos el procedimiento habitual de prueba de hipótesis paso a paso.

1. **Replantear el problema en función de hipótesis de investigación e hipótesis nula de las poblaciones.** Existen tres poblaciones.

Población 1: alcohólicos que reciben el tratamiento A.
Población 2: alcohólicos que reciben el tratamiento B.
Población 3: alcohólicos que reciben el tratamiento C.

La hipótesis nula establece que las tres poblaciones tienen la misma media. La hipótesis de investigación establece que no todas tienen la misma media.

2. **Determinar las características de la distribución comparativa.** La distribución comparativa en un análisis de varianza es siempre una distribución F. Calculamos sus grados de libertad del mismo modo que lo hemos venido haciendo hasta ahora. **La estimación intergrupal de la varianza es la cantidad de grupos menos 1.** Existen tres grupos, por lo tanto gl_{Entre} es 2. El gl_{Dentro} es la cantidad de observaciones de cada grupo menos uno. Hay 3 grados de libertad en el primer grupo (4 observaciones menos 1) y dos grados de libertad en cada uno de los otros grupos; por lo tanto, gl_{Dentro} es 7. Es decir, se trata de una distribución F para 2 y 7 grados de libertad.

$$S^2_{\text{Dentro}} = \frac{gl_1}{gl_1 + gl_2 + \cdots + gl_{\text{Últimos}}}\left(S_1^2\right) + \frac{gl_2}{gl_1 + gl_2 + \cdots + gl_{\text{Últimos}}}\left(S_2^2\right) + \cdots + \frac{gl_{\text{Últimos}}}{gl_1 + gl_2 + \cdots + gl_{\text{Últimos}}}\left(S^2_{\text{Últimos}}\right)$$

Calcular la S^2_{Entre} es un poco más complejo. Primero, calculamos la gran media general (que no es sólo la media de las medias). Para calcular la gran media, primero multiplicamos la media de cada grupo por la cantidad de observaciones de ese grupo, sumamos los resultados de todos los grupos y dividimos la suma por la cantidad total de observaciones. Se expresa bajo la fórmula,

$$GM = \frac{(M_1)(n_1) + (M_2)(n_2) + \cdots + (M_{\text{Últimos}})(n_{\text{Últimos}})}{n_1 + n_2 + \cdots + n_{\text{Últimos}}}$$

Después calculamos S^2_{Entre}: calculamos la desviación de la media de cada grupo con respecto a la gran media; elevamos la desviación al cuadrado; multiplicamos las desviaciones cuadráticos de cada grupo por la cantidad de observaciones del grupo; sumamos los resultados de todos los grupos y dividimos esa suma por los grados de libertad intergrupales (gl_{Entre} = cantidad de grupos menos 1). Se expresa bajo la fórmula,

$$S^2_{\text{Entre}} = \frac{(M_1 - GM)^2(n_1) + (M_2 - GM)^2(n_2) + \cdots + (M_{\text{Última}} - GM)^2(n_{\text{Último}})}{gl_{\text{Entre}}}$$

3. **Determinar el punto muestral de corte en la distribución comparativa, a partir del cual debería rechazarse la hipótesis nula.** Utilizando la tabla F del apéndice B (tabla B-3), buscamos en la columna correspondiente a 2 grados de libertad en el numerador y nos detenemos a los 7 grados de libertad del denominador. Utilizando el nivel 0,05 de significación (el número del medio), encontramos un punto de corte de 4,74.

4. **Determinar el valor muestral en la distribución comparativa.** Dado que la distribución comparativa es una distribución de razones F, este paso implica calcular la razón F de la muestra (utilizando el método estudiado en este capítulo). El numerador es la estimación intergrupal de varianza. Se basa en la desviación de la media del grupo de cada observación con respecto a la gran media. Por ejemplo, la media del grupo de la primera observación es 10, y la gran media es 7. La desviación es 3, y la desviación cuadrática es 9. Sumando las 10 desviaciones cuadráticas de este tipo obtenemos 66, que aparece en la tabla del análisis de varianza bajo la columna SC, en la fila "Intergrupal". Después, dividimos la suma de las desviaciones cuadráticas por los grados de libertad intergrupales (gl_{Entre}). El resultado, como lo indica la tabla del análisis de varianza, bajo CM, resulta ser 33. Es decir, 33 es el numerador, la estimación intergrupal de la varianza poblacional.

Tabla 12-3.
Análisis de varianza del estudio acerca de tratamientos de alcoholismo (datos ficticios).

Tratamiento A							Tratamiento B							Tratamiento C						
X	$X-GM$		$X-M$		$M-GM$		X	$X-GM$		$X-M$		$M-GM$		X	$X-GM$		$X-M$		$M-GM$	
	Desv	*Desv²*	*Desv*	*Desv²*	*Desv*	*Desv²*		*Desv*	*Desv²*	*Desv*	*Desv²*	*Desv*	*Desv²*		*Desv*	*Desv²*	*Desv*	*Desv²*	*Desv*	*Desv²*
8	1	1	−2	4	3	9	7	0	0	1	1	−1	1	6	−1	1	2	4	−3	9
13	6	36	3	9	3	9	3	−4	16	−3	9	−1	1	4	−3	9	0	0	−3	9
10	3	9	0	0	3	9	8	1	1	2	4	−1	1	2	−5	25	−2	4	−3	9
9	2	4	−1	1	3	9														
40		50		14		36	18		17		14		3	12		35		8		27
$M = 40/4 = 10$							$M = 18/3 = 6$							$M = 12/3 = 4$						

Nota: Desv = Desviación. Desv² = Desviación cuadrática

$GM = (40 + 18 + 12)/10 = 70/10 = 7$

$gl_{\text{Total}} = N - 1 = 10 - 1 = 9$

$gl_{\text{Dentro}} = gl_1 + gl_2 + \ldots + gl_{\text{Último}} = (4 - 1) + (3 - 1) + (3 - 1) = 3 + 2 + 2 = 7$

$gl_{\text{Entre}} = N_{\text{Grupos}} - 1 = 3 - 1 = 2$

F necesario para $gl = 2, 7$ al nivel $0{,}05 = 4{,}74$

$SC_{\text{Total}} = 50 + 17 + 35 = 102$

$SC_{\text{Dentro}} = 14 + 14 + 8 = 36$

$SC_{\text{Entre}} = 36 + 3 + 27 = 66$

TABLA DEL ANÁLISIS DE VARIANZA:

FUENTE	*SC*	*GL*	*CM*	*F*
Intergrupal	66	2	33	6,42
Intragrupal	36	7	5,14	
Total	102	9		

Conclusión: Se rechaza la hipótesis nula.

El denominador de la razón F es la estimación intragrupal de la varianza poblacional. Se basa en las desviaciones de cada observación *con respecto a la media de su grupo.* Por ejemplo, la primera observación es 8 y la media de su grupo es 10. Esto da una desviación de −2 y una desviación cuadrática de 4. Sumando las 10 desviaciones cuadráticas de este tipo, obtenemos 36. Dividimos 36 por los grados de libertad intragrupals, que son 7, y obtenemos 5,14.

La razón F, como siempre, es la estimación intergrupal dividida por la estimación intragrupal. El resultado es 6,42.

Al utilizar el método tratado en este capítulo, y al trabajar manualmente, conviene a esta altura calcular la suma de las desviaciones cuadráticas de cada observación con respecto a la gran media, que aparece en la línea correspondiente a "Total" de la tabla. De ese modo podemos controlar los cálculos aritméticos, ya que esta suma debería ser igual al total de las otras dos sumas de desviaciones cuadráticas. (En este caso, 66 más 36 es igual a 102).

5. **Comparar los valores obtenidos en los pasos 3 y 4 para decidir si se rechaza o no la hipótesis nula.** La razón F de 6,42 es más extrema que el punto de corte F de 4,74 correspondiente al nivel 0,05 de significación. Por lo tanto, el investigador puede rechazar la hipótesis nula. Si esta

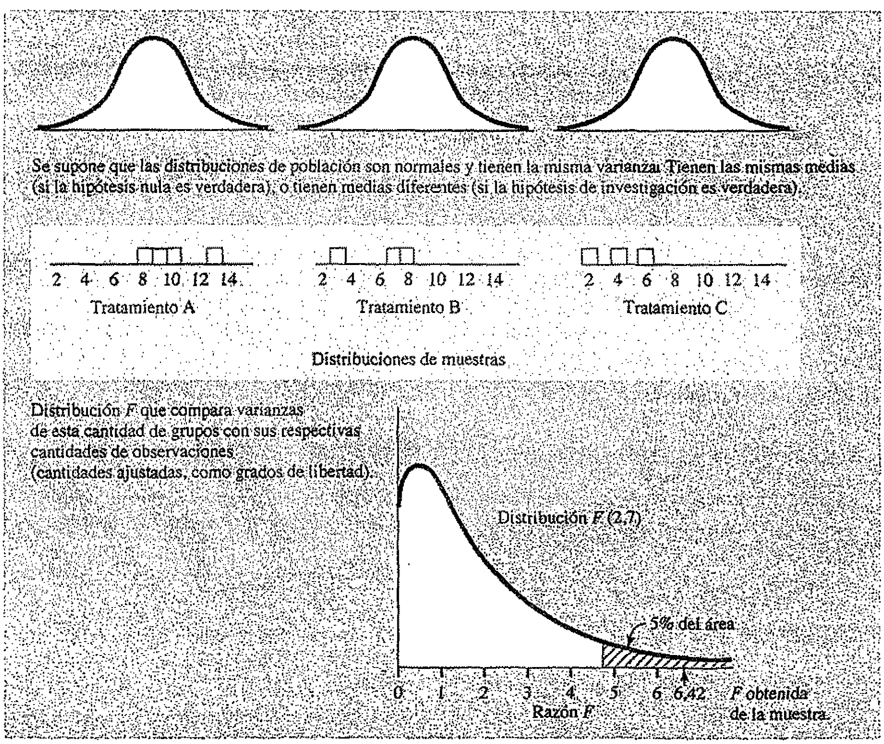


Figura 12-3.
Distribuciones relacionadas con el estudio ficticio acerca del tratamiento del alcoholismo.

fuera información real, el investigador podría concluir que los tres tratamientos tienen diferentes efectos en cuanto a la satisfacción de pacientes, como los suyos con respecto a sus tratamientos.

Otro Ejemplo

Ahora examinaremos información ficticia basada en resultados de un estudio real realizado por Clark et al. (1997). Los investigadores estudiaron tres grupos de pacientes: pacientes con pánico, pacientes con angustia generalizada y pacientes con fobia social. También incluyeron un grupo comparativo de personas que no eran pacientes. Como parte inicial del estudio, compararon los cuatro grupos sobre la base de varias medidas estándar. La tabla 12-4 se basa en los descubrimientos reales de los investigadores a través de las pruebas de ansiedad. (El patrón de los resultados es el mismo. Sin embargo, para que el ejemplo fuera simple, hemos utilizado muchos menos participantes y hemos transformado los valores individuales en números agradables, enteros y pequeños. Los resultados del estudio real se indican en la tabla 12-8, más adelante en este capítulo). La tabla 12-4 también presenta los cálculos principales y la tabla del análisis de varianza. La figura 12-4 representa gráficamente las distintas distribuciones relacionadas con el estudio. A continuación, analizamos el ejemplo siguiendo el procedimiento normal de prueba de hipótesis paso a paso.

Tabla 12-4.
Análisis de varianza de valores de ansiedad basado aproximadamente en Clark et al. (1997). (Datos ficticios).

No pacientes				Pacientes con pánico				Paciente con angustia generalizada				Paciente con fobia social			
Desviaciones cuadráticas				*Desviaciones cuadráticas*				*Desviaciones cuadráticas*				*Desviaciones cuadráticas*			
X	$X-GM$	$X-M$	$M-GM$	X	$X-GM$	$X-M$	$M-GM$	X	$X-GM$	$X-M$	$M-GM$	X	$X-GM$	$X-M$	$M-GM$
7	9	1	4	11	1	0	1	10	0	1	1	11	1	0	1
8	4	0	4	10	0	1	1	12	4	1	1	11	1	0	1
10	0	4	4	12	4	1	1					11	1	0	1
7	9	1	4												
32	22	6	16	33	5	2	3	22	4	2	2	33	3	0	3
$M = 32/4 = 8$				$M = 33/3 = 11$				$M = 22/2 = 11$				$M = 33/3 = 11$			

$GM = (32 + 33 + 22 + 33)/12 = 10$

$gl_{\text{Total}} = N - 1 = 12 - 1 = 11$

$gl_{\text{Dentro}} = gl_1 + gl_2 + \ldots + gl_{\text{Último}} = (4 - 1) + (3 - 1) + (2 - 1) + (3 - 1) = 3 + 2 + 1 + 2 = 8$

$gl_{\text{Entre}} = N_{\text{Grupos}} - 1 = 4 - 1 = 3$

F necesario para $gl = 3, 8$ al nivel 0,05= 4,07

$SC_{\text{Total}} = 22 + 5 + 4 + 3 = 34$

$SC_{\text{Dentro}} = 6 + 2 + 2 + 0 = 10$

$SC_{\text{Entre}} = 16 + 3 + 2 + 3 = 24$

ANALISIS DE VARIANZA:

Fuente	SC	gl	CM	F
Intergrupal	24	3	8	6.4
Intragrupal	10	8	1,25	
Total	34	11		

Conclusión: se rechaza la hipótesis nula.

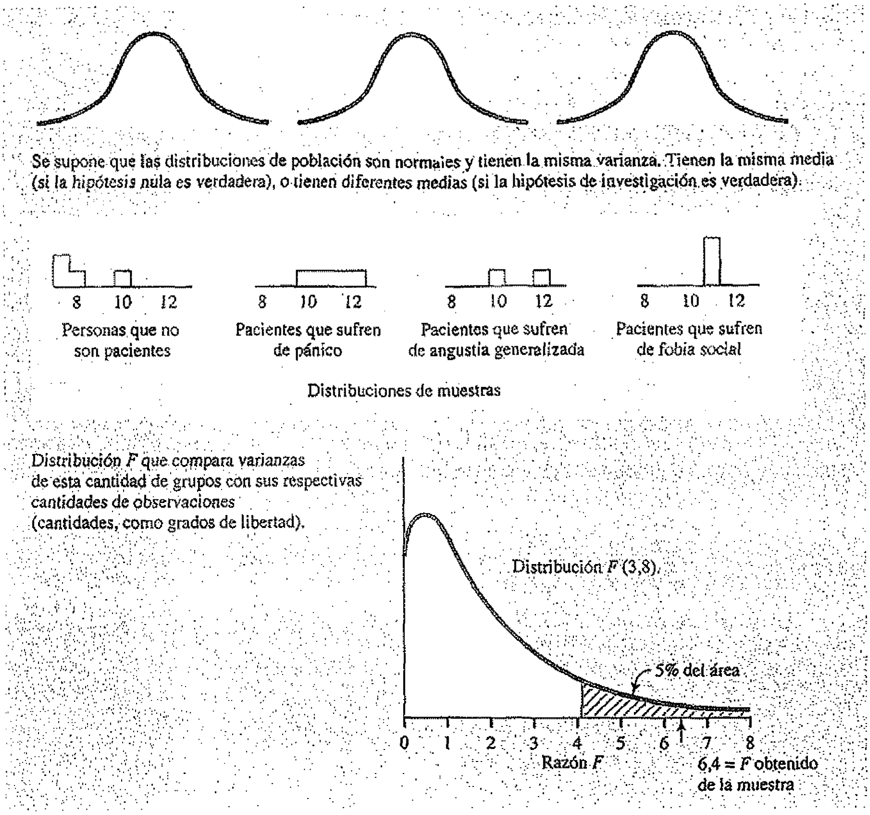


Figura 12-4.
Distribuciones relacionadas con el análisis de varianza de la información ficticia basado aproximadamente en Clark et al.

1. **Replantear el problema en función de hipótesis de investigación e hipótesis nula de las poblaciones.** Existen cuatro poblaciones:

Población 1: personas que no son pacientes.
Población 2: pacientes que sufren de pánico.
Población 3: pacientes que sufren de angustia generalizada.
Población 4: pacientes que sufren de fobia social.

La hipótesis nula establece que las cuatro poblaciones tienen la misma media en cuanto al nivel de ansiedad. La hipótesis de investigación establece que no todas tienen la misma media en cuanto al nivel de ansiedad.

2. ***Determinar las características de la distribución comparativa***. La distribución comparativa es una distribución F con 3 y 8 grados de libertad, tal como lo indica la figura 12-4.

3. **Determinar el punto muestral de corte en la distribución comparativa, a partir del cual debería rechazarse la hipótesis nula.** La tabla B-3 indica que con 3 grados de libertad en el numerador y 8 en el denominador, el punto de corte al nivel 0,05 es de 4,07 (*véase* figura 12-4).

4. **Determinar el valor muestral en la distribución comparativa.** La tabla 12-4 indica los cálculos de la razón F de la muestra, utilizando el método del modelo estructural tratado en este capítulo. El numerador es la estimación intergrupal de la varianza que se basa en la desviación de la media del grupo de cada observación con respecto a la gran media, es decir, 8. El denominador es la estimación intragrupal de la varianza que se basa en la desviación de cada observación con respecto a la media de su grupo, es decir, 1,25. Por lo tanto, la razón F es 6,4 (es decir, 8/1,25 = 6,4).

5. **Comparar los valores obtenidos en los pasos 3 y 4 para decidir si se rechaza o no la hipótesis nula.** El F muestral igual a 6,4 es mayor al punto de corte F de 4,07. Por lo tanto, podemos rechazar la hipótesis nula y concluir que los cuatro grupos tienen diferentes niveles de ansiedad (*véase* figura 12-4).

RESUMEN DE LOS PROCEDIMIENTOS PARA EL CÁLCULO DEL ANÁLISIS DE VARIANZA UTILIZANDO EL MODELO ESTRUCTURAL

La mitad superior de la tabla 12-5 resume los pasos para la realización de un análisis de varianza utilizando el método tratado en este capítulo. La mitad inferior muestra una tabla del análisis de varianza con los símbolos de todas las partes insertados en cada sección, donde, usualmente, irían los números. Más abajo hay una tabla del análisis de varianza del mismo estilo, con las distintas fórmulas donde generalmente irían los números. Es importante tener en cuenta que la única diferencia con lo que hicimos en el capítulo 11 se encuentra en el paso 4, en los puntos que van del b al g. No debemos olvidar tampoco que este es el método que se utiliza cuando los grupos tienen tamaños desiguales.

El apéndice del capítulo nos proporciona las fórmulas de cálculo para el análisis de varianza basado en el método del modelo estructural (y por lo tanto adecuadas para utilizar con muestras de tamaños desiguales). Esas fórmulas serán útiles cuando no se pueda utilizar una computadora y cuando sea necesario calcular un análisis de varianza de este tipo para un estudio real con una gran cantidad de participantes. Sin embargo, para aprender la lógica del análisis de varianza recomendamos insistentemente que los ejercicios se realicen utilizando las fórmulas de definición y los procedimientos indicados en la tabla 12-5.

COMPARACIONES MÚLTIPLES

Rechazar la hipótesis nula en un análisis de varianza implica que las medias poblacionales no son todas iguales. Lo que no queda claro, sin embargo, es cuáles son las medias poblacionales que difieren entre sí. Por ejemplo, en el estudio acerca de los antecedentes delictivos, los miembros del jurado que formaban el grupo al que se le informó la existencia de antecedentes delictivos fueron los que asignaron el mayor nivel de culpabilidad ($M = 8$); los miembros del jurado que no recibieron información al respecto fueron los segundos en cuanto al nivel de culpabilidad asignado ($M = 5$), y los miembros del jurado a los que se informó que el acusado no tenía antecedentes delictivos fueron los que asignaron el nivel más bajo de culpabilidad ($M = 4$). A partir de los resultados del análisis de varianza, concluimos que las verdaderas medias de las tres poblaciones que representaban estos grupos no eran todas iguales. Sin embargo, no sabemos qué medias de qué poblaciones en particular son significativamente diferentes entre sí. Ni siquiera existe garantía de

Tabla 12-5.
Pasos, símbolos y fórmulas para calcular un análisis de varianza utilizando el método del modelo estructural (grupos de tamaños iguales o desiguales).

Pasos de la prueba de hipótesis

1. Replantear el problema en función de hipótesis de investigación e hipótesis nula de las poblaciones.
2. Determinar las características de la distribución comparativa.
   a) La distribución comparativa será una distribución *F*.
   b) Los grados de libertad del numerador son la cantidad de grupos menos 1; $gl_{\text{Entre}} = N_{\text{Grupos}} - 1$.
   c) Los grados de libertad del denominador son la suma de los grados de libertad de cada grupo (la cantidad de observaciones de cada grupo menos 1): $gl_{\text{Dentro}} = gl_1 + gl_2 + \ldots + gl_{\text{Último}}$.
   d) Controlar la exactitud de los cálculos asegurándose de que gl_{Dentro} más gl_{Entre} suman gl_{Total} (que es la cantidad total de casos menos 1).
3. Determinar el punto muestral de corte en la distribución comparativa, a partir del cual debería rechazarse la hipótesis nula.
   a) Determinar el nivel de significación deseado.
   b) Buscar el punto de corte correspondiente a la tabla *F*.
4. Determinar el valor muestral en la distribución comparativa (será un razón *F*).
   a) Calcular la media de cada grupo y la gran media de todas las observaciones.
   b) Calcular las siguientes desviaciones para cada observación:
      i) La desviación con respecto a la gran media $(X - GM)$.
      ii) La desviación con respecto a la media de su grupo $(X - M)$.
      iii) La desviación de la media de su grupo con respecto a la gran media $(M - GM)$.
   c) Elevar al cuadrado cada una de esas desviaciones.
   d) Calcular las sumas de cada uno de estos tres tipos de desviaciones cuadráticas (SC_{Total}, SC_{Dentro} y SC_{Entre}).
   e) Controlar la exactitud de los cálculos asegurándose de que $SC_{\text{Dentro}} + SC_{\text{Entre}} = SC_{\text{Total}}$.
   f) Calcular la estimación intergrupal de varianza: $SC_{\text{Entre}}/gl_{\text{Entre}}$.
   g) Calcular la estimación intragrupal de varianza: $SC_{\text{Dentro}}/gl_{\text{Dentro}}$.
   h) Calcular la razón *F*: $F = S^2_{\text{Entre}}/S^2_{\text{Dentro}}$ ó $F = CM_{\text{Entre}}/CM_{\text{Dentro}}$.
5. Comparar los resultados obtenidos en los pasos 3 y 4 para decidir si se rechaza o no la hipótesis nula.

Símbolos correspondientes a cada parte de un análisis de varianza

Fuente	SC	gl	CM	F
Intergrupal	SC_{Entre}	gl_{Entre}	CM_{Entre} (ó S^2_{Entre})	F
Intragrupal	SS_{Dentro}	gl_{Dentro}	CM_{Dentro} (ó S^2_{Dentro})	
Total	SC_{Total}	gl_{Total}		

Fórmulas correspondientes a cada parte de un análisis de varianza

Fuente	SC	gl	CM	F
Intergrupal	$\Sigma(M - GM)^2$	$N_{\text{Grupos}} - 1$	$SC_{\text{Entre}}/gl_{\text{Entre}}$	$CM_{\text{Entre}}/CM_{\text{Dentro}}$
Intragrupal	$\Sigma(X - M)^2$	$gl_1 + gl_2 + \ldots + gl_{\text{Último}}$		$SC_{\text{Dentro}}/gl_{\text{Dentro}}$
Total	$\Sigma(X - GM)^2$	$N - 1$		

que los dos grupos más extremadamente diferentes (el grupo que recibió información de antecedentes delictivos y el grupo al que se le informó que no existían antecedentes delictivos) representen poblaciones con medias diferentes y, ciertamente, no queda claro si la media de la población correspondiente al grupo que no recibió información al respecto es diferente de cualquiera de las medias de las poblaciones representadas por los otros dos grupos.

Cuando se determina cuáles son las medias que difieren entre sí se dice que se realizan **comparaciones múltiples**, porque frecuentemente se comparan varios pares de medias. Las comparaciones múltiples son un tema complejo muy tratado en cursos de estadística en psicología de nivel intermedio. Además, es un tema controvertido.

Existe un punto en el que casi todo el mundo está de acuerdo. Por lo general **no** es suficiente calcular simplemente una serie de pruebas *t*, una para cada posible par de medias, ya que si no se aplican modificaciones, es muy probable que este tipo de procedimiento arroje lo que aparentemente es un resultado significativo. Por ejemplo, con tres grupos existirían tres pruebas *t* posibles (el grupo 1 comparado con el 2, el 2 con el 3 y el 1 con el 3). Supongamos que utilizamos el nivel 0,05, de forma tal que cada una de las tres pruebas *t* posibles tengan una probabilidad 0,05 de resultar significativas equivocadamente. La probabilidad de que al menos una de las pruebas de la serie de tres pruebas *t* resulte significativa por equivocación, es aproximadamente del 15%. Con cuatro grupos, podría haber seis comparaciones. Lo cual significa que si usáramos el nivel 0,05 para cada prueba, tendríamos un riesgo total de casi el 30% de que, al menos, una resulte significativa sólo por casualidad.[2] Más aún, un investigador puede necesitar hacer comparaciones adicionales que no comparan simplemente a un grupo con otro; por ejemplo, se puede comparar el promedio de tres grupos con un cuarto grupo (tal vez los primeros tres son diferentes tipos de grupos experimentales y el cuarto es el grupo de control). La cantidad de comparaciones, aun con una cantidad bastante pequeña de grupos, puede ser considerablemente grande.

La controversia surge cuando los estadísticos intentan ponerse de acuerdo acerca de la mejor alternativa para no realizar simplemente un puñado de pruebas *t*. Las soluciones disponibles dependen, en parte, de la situación.

Comparaciones planificadas

Existe un tipo de situación que se presenta cuando el investigador ha decidido previamente observar unas pocas comparaciones en particular que están directamente relacionadas con la teoría o con alguna aplicación práctica. A esto se lo denomina **comparaciones planificadas** (o, a veces, **comparaciones a priori o contrastes planificados**), porque han sido planificadas previamente a la realización del estudio. (Estas comparaciones también son lo que habitualmente se denominan "contrastes lineales"). Analicemos nuevamente el ejemplo del estudio acerca de los antecedentes delictivos. El investigador podría decidir previamente que las únicas comparaciones de interés son a) el grupo que recibió los antecedentes delictivos con el grupo al que se informó que no existían antecedentes delictivos y b) el grupo que recibió los antecedentes delictivos con el grupo que no recibió información al respecto.

[2] En realidad, la probabilidad de obtener al menos un resultado significativo por casualidad, de tres, al nivel 0,05, es 0,143; y de obtener al menos uno de seis, es de 0,265. La fórmula para tres pruebas es $1 - (1 - \alpha)(1 - \alpha)(1 - \alpha)$, en donde α representa el nivel de significación. Además, es evidente que toda esta cuestión está muy relacionada con el tema de "demasiadas pruebas *t*" que analizamos en el capítulo 10. La cuestión planteada en ese capítulo se refería a dos grupos con diferencias en varias variables. En este caso, estamos hablando de diferencias entre varios grupos en cuanto a una variable. Por supuesto, algunas veces se presentan a la vez varios grupos y varias variables.

Un método ampliamente utilizado para analizar las comparaciones planificadas es el **procedimiento Bonferroni** (también llamado la "prueba de Dunn"), que se basa en la idea de utilizar un nivel de significación más exigente para cada comparación. El resultado es que la probabilidad total de que cualquiera de las comparaciones resulte significativa por error sigue siendo razonablemente baja. Por ejemplo, si cada una de dos comparaciones planificadas utilizara el nivel 0,025 de significación, la probabilidad total de que cualquiera de ellas resulte erróneamente significativa aún sería de menos del 0,05. Con tres comparaciones planificadas, podríamos utilizar el nivel 0,017 (3 por 0,017 es igual a 0,05). A veces, los investigadores que tienen dos o tres contrastes planificados simplemente utilizan el nivel 0,01 para cada uno, ya que se trata de un caso con un punto de corte con el que están familiarizados y que resulta fácil de encontrar en las tablas.

Comparaciones *post hoc*

Una situación muy diferente a la de las comparaciones planificadas es aquella en la que, después de haberse realizado el estudio, el investigador simplemente busca entre los resultados tratando de descubrir cuáles son los grupos que difieren entre sí. A estas comparaciones se las denomina **comparaciones *post hoc*** (o **comparaciones a posteriori**), porque no se planifican previamente.

Cuando se realizan comparaciones *post hoc*, se deben tener en cuenta todas las posibles comparaciones para calcular la probabilidad total de que cualquiera de ellas resulte significativa. Por ese motivo, utilizar el procedimiento Bonferroni para las comparaciones *post hoc* es seguro, pero cualquiera de las comparaciones presenta muy baja potencia. El nivel 0,05 se divide en tantas partes que, obtener alguna comparación significativa, sería extremadamente raro. Por lo tanto, los estadísticos han desarrollado una variedad de procedimientos para utilizar en estas búsquedas exploratorias. Los procedimientos mencionados intentan mantener el alfa general a un nivel cercano al 0,05, sin reducir de manera demasiado drástica la potencia. Algunos de estos procedimientos aparecen en las publicaciones descriptos por los nombres de aquellos que los desarrollaron; los métodos Scheffé, Tukey, Neuman-Keuls y Duncan son los más utilizados. Aún se discute qué procedimiento es más conveniente en distintas condiciones. Las distintas posibilidades y controversias acerca de este tema son tratadas en cursos de estadística de nivel medio.

Efecto producido por los diferentes métodos de realización de las comparaciones múltiples

Las comparaciones planificadas casi siempre tienen más potencia que las comparaciones *post hoc*. Además, cuantas menos comparaciones se planifican, mayor es la potencia. A veces, un investigador está, antes de realizar el estudio, en una posición que le permite disminuir la cantidad de comparaciones. Si antes de realizar el estudio existen sólidos fundamentos para realizar sólo determinadas comparaciones, ésto puede aumentar mucho la potencia del estudio. Sin embargo, según nuestra experiencia, en la mayoría de las situaciones reales de investigación el resultado es bastante parecido, cualquiera sea el procedimiento utilizado. Si un resultado es tan ajustado como para que el método sí influya, en cualquier caso es probable que se deba tomar el resultado con verdadera precaución. Lo que realmente importa es que el investigador utilice un método apropiado para enfrentar los riesgos de las comparaciones múltiples, y no que simplemente realice series de 50 pruebas *t* como si cada una fuera un mundo aparte.

SUPUESTOS DEL ANÁLISIS DE VARIANZA CON MUESTRAS DE TAMAÑOS DESIGUALES

En el capítulo 11 analizamos los supuestos del análisis de varianza. Sin embargo, cuando los tamaños de los grupos no son aproximadamente iguales, el análisis de varianza es mucho más sensible al incumplimiento del supuesto que establece que las varianzas deben ser iguales. De hecho, *con tamaños de muestras desiguales el análisis de varianza se torna sospechoso cuando las estimaciones de las varianzas poblacionales más extremadamente diferentes (entre aquellas de distintos grupos) son incluso tan diferentes como 1 1/2 vez una de la otra.*

TAMAÑO DEL EFECTO Y POTENCIA[3]

En el capítulo 11 presentamos el concepto de f, un indicador del tamaño del efecto para el análisis de varianza. f sigue siendo un tamaño del efecto completamente adecuado, ya sea que el análisis de varianza se calcule utilizando el método del capítulo 11 ó el método del modelo estructural descripto en este capítulo. Sin embargo, una desventaja de f es que su significado no es fácil de captar intuitivamente incluso para los investigadores más experimentados.

No obstante, existe otra medida comúnmente utilizada del tamaño del efecto en el análisis de varianza, cualquiera sea el método aplicado. (No podíamos presentar antes esta medida alternativa porque se la calcula utilizando elementos del método del modelo estructural). Primero describiremos la medida y la forma en que se calcula; luego veremos cómo se interpreta y por qué proporciona una indicación más intuitivamente, significativa que f acerca del tamaño del efecto.

Proporción de la varianza explicada

Un indicador del tamaño del efecto en el análisis de varianza, además de f, es la **proporción de la varianza explicada (R^2).** Vale la pena observar que utilizamos el mismo símbolo, R^2, para la proporción de la varianza explicada en el análisis de varianza que para la proporción de la varianza explicada en la correlación y regresión múltiples (*véase* capítulo 4). En ambos casos, la proporción de la varianza explicada describe hasta qué punto la variación en la variable dependiente puede ser explicada (predicha o justificada) por la variable independiente. En el análisis de varianza, la variable independiente se refiere al grupo al que pertenece una persona. Por lo tanto, el grado en el cual la variación en la variable dependiente es explicada por la variable independiente, es el grado en el cual el valor observado particular de una persona está relacionado o determinado por el grupo al que pertenece dicha persona. (En el capítulo 16 seguiremos hablando sobre los muchos vínculos entre el análisis de varianza y la correlación y regresión múltiples).

Para ser precisos, R^2 es la proporción total de la variación de las observaciones con respecto a la gran media que está explicada por la variación entre las medias de los grupos. Se calcula utilizando las sumas de los cuadrados, y es igual a la suma de cuadrados intergrupal (SC_{Entre}) dividida por la suma de cuadrados total (SC_{Total}). Se expresa bajo la fórmula,

$$R^2 = \frac{SC_{\text{Entre}}}{SC_{\text{Total}}} \tag{12-4}$$

[3] Al redactar esta sección suponemos que el alumno ya ha completado el capítulo 4.

Analicemos una vez más el estudio referido a los antecedentes delictivos. En ese ejemplo, la suma de las desviaciones cuadráticas de las observaciones con respecto a la gran media era 107,33, y la suma de las desviaciones cuadráticas de las medias de los grupos de las observaciones con respecto a la gran media era 43,44. Por lo tanto, la proporción de la variación total explicada por la variación entre los grupos es 43,44/107,33 ó 40%. Se expresa bajo la fórmula,

$$R^2 = \frac{SC_{\text{Entre}}}{SC_{\text{Total}}} = \frac{43{,}44}{107{,}33} = 0{,}40$$

Qué sucede si, como ocurre con frecuencia en los estudios publicados, las sumas de los cuadrados no están disponibles. También es posible calcular R^2 directamente a partir de F y de los grados de libertad. La fórmula es la siguiente:

$$R^2 = \frac{(F)(gl_{\text{Entre}})}{(F)(gl_{\text{Entre}}) + gl_{\text{Dentro}}} \tag{12-5}$$

Por ejemplo, en el estudio acerca de los antecedentes delictivos,

$$R^2 = \frac{(F)(gl_{\text{Entre}})}{(F)(gl_{\text{Entre}}) + gl_{\text{Dentro}}} = \frac{(4{,}07)(2)}{(4{,}07)(2)+12}$$

$$= \frac{8{,}14}{8{,}14+12} = \frac{8{,}14}{20{,}14} = 0{,}40$$

Interpretación de la proporción de la varianza explicada

La proporción de la varianza explicada es una medida útil del tamaño del efecto, ya que lleva implícito su significado directo en su propio nombre y, además, los investigadores están familiarizados con R^2 por su utilización en la regresión. Finalmente, la proporción de varianza explicada es útil porque su raíz cuadrada, R, es una especie de coeficiente de correlación con el que la mayoría de los investigadores están muy familiarizados (*véase* capítulos 3 y 4).

R^2 es una proporción de dos números positivos ($SC_{Entre}/SC_{\text{Total}}$), por lo tanto, tiene un mínimo de 0 y un máximo de 1. Sin embargo, en la práctica es raro que, en un análisis de varianza, el R^2 llegue siquiera a 0,50. La tabla 12-6 indica las reglas de Cohen (1988) para el tamaño del efecto correspondiente a R^2, junto con valores comparables correspondientes a f y R. (Con estos valores aún podemos utilizar las tablas de potencia y de tamaño de muestra del capítulo 11, cuando trabajamos con R o R^2 en lugar de f). Es importante observar que aun un gran tamaño del efecto para R^2 es de tan sólo aproximadamente el 14% y que, además, en los tres niveles de tamaño del efecto indicados, f y R son bastante similares.[4] Sin embargo, a niveles muy altos no lo serían tanto; por ejemplo, una f de 1,0 corresponde a una R de 0,71. Sin embargo, a los niveles de tamaño del efecto comunes en el análisis de varianza, puede ser bastante útil considerar a f como aproximadamente correspondiente a un coeficiente de correlación.

[4] La relación exacta entre R^2 y f es $R^2 = f^2/(1+f^2)$ y $f = \sqrt{R^2/(1-R^2)}$. Sin embargo, si intentamos calcular una a partir de la otra utilizando información tomada de un estudio real, los resultados no coincidirán exactamente con lo que obtenemos cuando calculamos cada una directamente. Esto ocurre porque f se basa en desvíos estándar de población estimados, y R^2 es una descripción directa de información de la muestra.

Tabla 12-6.
Reglas de Cohen para tamaños del efecto en un análisis de varianza de un criterio.

	Tamaño del efecto		
	Pequeño	*Mediano*	*Grande*
f	0,10	0,25	0,40
R	0,10	0,24	0,37
R^2	0,01	0,06	0,14

También debemos saber que otro nombre común para esta medida del tamaño del efecto (además de R^2) es η^2, la letra griega **eta al cuadrado**; η^2 también se conoce como **razón de correlación.**

CONTROVERSIAS, LIMITACIONES Y DESARROLLOS RECIENTES

El análisis de varianza se utiliza comúnmente en situaciones en las que se comparan tres o más grupos. (Si se comparan dos grupos, se puede utilizar una prueba t). Sin embargo, Rosnow y Rosenthal (1989) sostienen que esas pruebas "difusas" o "colectivas" no son muy útiles. Sostienen que, en casi todos los casos cuando probamos la diferencia general entre tres o más grupos, "hemos probado un aspecto en el cual casi con seguridad no estamos interesados", (p. 1281) ¿En qué aspectos estamos **efectivamente** interesados? Concretamente en comparaciones entre pares específicos, ya sea entre dos grupos o bien entre uno y una combinación de grupos.

Rosnow y Rosenthal defienden la idea de que, al calcular el análisis de varianza, deberíamos analizar sólo comparaciones planificadas. Esas comparaciones planificadas deberían reemplazar por completo la prueba F general (es decir, la difusa o colectiva prueba F) para decidir si se puede rechazar la hipótesis de la inexistencia de diferencias entre las medias poblacionales. Tradicionalmente, las comparaciones planificadas, cuando se utilizan son un complemento de la prueba F general; por eso, ésta sería una idea bastante revolucionaria.

Analicemos un ejemplo. Orbach et al. (1997) realizaron un estudio que utilizamos en el capítulo 2, como ejemplo de cómo se describen las medias y desvíos estándar en las publicaciones científicas. Los investigadores compararon un grupo de pacientes suicidas de un hospital psiquiátrico (individuos que habían cometido serios intentos de suicidio), pacientes no suicidas de un hospital psiquiátrico con diagnósticos similares y un grupo de control de voluntarios tomados de la comunidad. El objetivo del estudio era probar la teoría de que los individuos suicidas tienen mayor tolerancia al dolor físico. La idea era que su mayor umbral de dolor les facilita la realización de los dolorosos actos que implica un suicidio. Los investigadores realizaron a los tres grupos pruebas estándar para medir el umbral de dolor, otras pruebas sensoriales y una variedad de cuestionarios. La siguiente es la descripción del análisis :

> Para analizar la hipótesis del estudio realizamos una serie de dos contrastes lineales para cada medida de dolor [...] El primer contraste lineal, **contraste de capacidad suicida**, comparaba al grupo suicida con los otros dos grupos no suicidas (internos psiquiátricos y participantes de control). El segundo contraste comparaba a los dos grupos no suicidas [...] No calculamos una F colectiva previa porque realizamos comparaciones de grupos previamente planificadas, las cua-

les probaban la hipótesis del estudio. Debido a que se necesitaban comparaciones múltiples, el alfa se estableció en 0,01 para evitar el error Tipo I.

El contraste de capacidad suicida fue significativo en cuanto al umbral de sensación térmica, $F(1, 95) = 21{,}64$, $p < 0{,}01$; umbral de dolor, $F(1, 95) = 23{,}65$, $p < 0{,}01$; tolerancia al dolor $F(1, 95) = 6{,}55$, $p < 0{,}01$, y tolerancia máxima $F(1, 95) = 16{,}05$. No se encontró diferencia significativa entre el grupo suicida y los grupos no suicidas en la medida estimada de tamaño. Un análisis de las medias de la tabla 1 [*véase* tabla 2-6] sostiene nuestra hipótesis principal: los participantes suicidas, como se esperaba, presentaban umbrales más altos de dolor y sensibilidad, mayor tolerancia al dolor, y tenían más posibilidades que los otros internos y el grupo de control de tolerar la temperatura máxima administrada. Es interesante observar que la segunda serie de contrastes no reveló diferencias significativas entre los internos con problemas psiquiátricos y los participantes de control, en cuanto a ninguna de las cinco medidas de dolor. (p. 648)

El estudio de Orbach et al. ejemplifica el consejo de Rosnow y Rosenthal de utilizar comparaciones planificadas en lugar de un análisis de varianza general. Pero este método todavía no ha sido adoptado en forma generalizada, y aún es controvertido. La principal preocupación es muy similar a la tratada en el capítulo 7 acerca de las pruebas de una y dos colas. Si adoptamos las comparaciones planificadas altamente dirigidas, que recomiendan Rosnow y Rosenthal, perdemos la posibilidad de encontrar diferencias inesperadas que no se planificaron inicialmente.

ANÁLISIS DE VARIANZA CON EL MODELO ESTRUCTURAL Y COMPARACIONES MÚLTIPLES SEGÚN SE DESCRIBEN EN LAS PUBLICACIONES CIENTÍFICAS

En el capítulo 11, vimos que los resultados de los análisis de varianza se describen generalmente en las publicaciones científicas a través del *F* y la significación. La mayoría de los programas para computadoras proporcionan usualmente una tabla del análisis de varianza del estilo que presentamos en este capítulo. Más aún, cuando los investigadores estaban menos familiarizados con los detalles acerca de la forma en que se realizaba el análisis de varianza, era común que las publicaciones científicas incluyeran una versión abreviada de una tabla del análisis de varianza; por eso, algunas veces se pueden llegar a ver esas tablas en las publicaciones más antiguas. Actualmente, la utilización y comprensión del análisis de varianza es tan generalizada que es raro encontrar esas tablas en las publicaciones científicas.

Existe un aspecto que no hemos tenido en cuenta hasta ahora en cuanto a la forma en que el análisis de varianza aparece en las publicaciones científicas, y que depende del material tratado en este capítulo. En líneas generales, al informar los resultados de cualquier análisis de varianza los investigadores también darán los resultados de las comparaciones múltiples. Algunas veces las comparaciones se describen en el texto de la publicación. Los resultados podrían mencionar que "las comparaciones planificadas se realizaron entre los grupos A y C y entre los grupos C y D. Ambas resultaron significativas". Anteriormente, en el estudio de Orbach et al. acerca de pacientes suicidas y tolerancia al dolor, vimos un ejemplo de este modo de informar los resultados.

En el caso de las comparaciones *post hoc*, los investigadores utilizan con frecuencia un procedimiento en el que agregan pequeñas letras al lado de las medias en las tablas. Comúnmente, las medias con la misma letra no son significativamente diferentes entre sí, y las que tienen diferentes letras sí lo son. Por ejemplo, la tabla 12-7 presenta los resultados reales de las medidas de experiencias amorosas en el estudio de Hazan y Shaver (1987) (nuestro primer ejemplo en el capítulo 11). Analicemos la primera fila (los resultados sobre la felicidad). Los grupos evasivos y ansiosos-ambivalentes no son significativamente diferentes entre sí, dado que tienen la misma letra (*a*). Pero ambos son significativamente diferentes en cuanto a felicidad, comparados con el grupo de individuos seguros que tiene una letra diferente (*b*). En la fila correspondiente a los celos, sin embargo, los tres grupos difieren entre sí.

Como segundo ejemplo, se reproduce la tabla 12-8 del estudio de Clark et al. (1997) que utilizamos anteriormente como ejemplo en este capítulo (con números basados aproximadamente en los resultados reales que aquí aparecen).

Al leer los resultados de comparaciones *post hoc*, veremos que se nombran muchos procedimientos diferentes, incluyendo "Neuman-Keuls", "HSD de Tukey", "rango múltiple de Duncan" y "Scheffé". Pero como ya dijimos, estos procedimientos son sólo diversos métodos para probar diferencias, realizados de distintas formas para intentar asegurar que la probabilidad de que cualquiera de ellas resulte significativa por error no sea inaceptablemente alta.

Por ejemplo, Miller (1997) pidió a 147 alumnas mujeres que miraran diapositivas de artículos de revistas que incluían, entre otras cosas, fotos de hombres atractivos. Mientras observaban los artículos, se midió su excitación física (potencia conductora de la piel); después de mirar las diapositivas se les pidió que calificaran el atractivo y cuánto les gustaría conocer a cada persona que aparecía en los artículos. Como parte de su análisis, Miller comparó los resultados de mujeres que no estaban relacionándose con nadie, mujeres que mantenían relaciones casuales y mujeres que mantenían relaciones exclusivas. La tabla 12-9 indica los resultados del estudio de Miller. Es interesante observar la mención a la prueba de rango múltiple de Duncan.

Tabla 12-7.
Medias de subescalas de amor correspondientes a los tres estilos de vinculación (muestra tomada de un periódico).

Nombre escala	Evasivo	Ansioso-ambivalente	Seguro	F (2, 571)
Felicidad	$3{,}19_a$	$3{,}31_a$	$3{,}51_b$	14,21***
Amistad	$3{,}18_a$	$3{,}19_a$	$3{,}50_b$	22,96***
Confianza	$3{,}11_a$	$3{,}13_a$	$3{,}43_b$	16,21***
Temor al acercamiento	$2{,}30_a$	$2{,}15_a$	$1{,}88_b$	22,65***
Aceptación	$2{,}86_a$	$3{,}03_b$	$3{,}01_b$	4,66**
Extremos emocionales	$2{,}75_a$	$3{,}05_b$	$2{,}36_c$	27,54***
Celos	$2{,}57_a$	$2{,}88_b$	$2{,}17_c$	43,91***
Preocupación obsesiva	$3{,}01_a$	$3{,}29_b$	$3{,}01_a$	9,47***
Atracción sexual	$3{,}27_a$	$3{,}43_b$	$3{,}27_a$	4,08*
Deseo de unión	$2{,}81_a$	$3{,}25_b$	$2{,}69_a$	22,67***
Deseo de reciprocidad	$3{,}24_a$	$3{,}55_b$	$3{,}22_a$	14,90***
Amor a primera vista	$2{,}91_a$	$3{,}17_b$	$2{,}97_a$	6,00**

Nota: Dentro de cada fila, las medias con diferentes subíndices difieren según el nivel 0,05 de significación, de acuerdo con la prueba de Scheffé.

*p < 0,05; **p < 0,01; ***p < 0,001.

Fuente: Hazan, C. & Shaver, P. (1987), tab. 3. "Amor romántico conceptuado como un proceso de vinculación". *Revista de Psicología Social y de la Personalidad [Journal of Personality and Social Psychology]*, 52, 511–524. Copyright, 1987, por la Asociación Americana de Psicología. Reimpreso con autorización.

Tabla 12-8.
Estudio 2: medias y desviaciones estándar correspondientes a las características de los participantes.

	M (y SD) por Grupo			
Variable	*Pánico* (n = 45)	*Angustia Generalizada* (n = 33)	*Fobia Social* (n = 73)	*No pacientes* (n = 45)
Edad	$33{,}0_a$ (7,1)	$40{,}1_b$ (9,6)	$34{,}9_a$ (8,9)	$33{,}0_a$ (6,9)
STAI	$48{,}8_b$ (12,1)	$49{,}5_b$ (9,5)	$46{,}4_b$ (10,0)	$29{,}2_a$ (5,4)
Inventario de depresión de Beck	$15{,}3_{b,c}$ (8,6)	$18{,}3_c$ (10,2)	$12{,}8_b$ (7,8)	$2{,}1_a$ (2,2)
VAS–Ansiedad	$23{,}0_b$ (18,6)	$28{,}8_b$ (22,1)	$25{,}0_b$ (18,2)	$5{,}6_a$ (9,4)
VAS–Depresión	$21{,}8_b$ (21,1)	$29{,}4_b$ (21,1)	$24{,}7_b$ (18,5)	$8{,}2_a$ (11,3)
VAS–Felicidad	$53{,}1_b$ (16,3)	$55{,}7_b$ (17,0)	$53{,}0_b$ (17,1)	$74{,}5_a$ (15,1)

Nota: Las medias con diferentes subíndices difieren significativamente ($p < 0{,}01$): STAI (*State-Trait Anxiety Invenory, State Subscale*, Inventario de ansiedad o estado, subescala de estado); VAS (*Visual analogue scale*, Escala análoga visual).

Fuente: Clark, D. M., et al. (1997), tab. 3. "Malas interpretaciones de sensaciones corporales en enfermos con pánico". *Revista de psicología clínica y consultiva [Journal of Consulting and Clinical Psychology]*, 65, 203–213. Copyright, 1997, por la Asociación Americana de Psicología. Reimpreso con autorización.

Tabla 12-9.
Efectos del tipo de relaciones.

	Tipo de relación		
Medida dependiente	*Ninguna*	*Casual*	*Exclusiva*
Potencia conductora de la piel	$19{,}5_b$	$19{,}1_b$	$15{,}8_a$
Deseo de conocer al objetivo	$14{,}6_b$	$15{,}3_b$	$11{,}2_a$
Atractivo físico percibido del objetivo	$15{,}6_b$	$17{,}1_b$	$13{,}8_a$

Nota: Los números más altos reflejan mayor excitación, deseo de conocer al objetivo y atractivo percibido; para los dos últimos ítems, el rango posible era 1–19. Dentro de cada fila, las medias con diferentes subíndices difieren significativamente ($p < 0{,}05$), prueba de rango múltiple de Duncan.
Fuente: Miller, R. S. (1997), tab. 4. "Desatento y satisfecho: compromiso en la relación y atención a alternativas". *Revista sobre Psicología Social y de Personalidad [Journal of Personality and Social Psychology]*, 73, 758–766. Copyright, 1997, por la Asociación Americana de Psicología. Reimpreso con autorización.

Resumen

Existe un enfoque alternativo del análisis de varianza, que se denomina modelo estructural. En el enfoque del modelo estructural, la desviación de cada observación con respecto a la gran media se divide en dos partes: a) la diferencia entre el valor observado y la media de su grupo y b) la diferencia entre la media de su grupo y la gran media. Estas desviaciones, al elevarse al cuadrado y sumarse y dividirse por los grados de libertad adecuados, dan las mismas estimaciones de las varianzas intragrupales e intergrupales que las obtenidas utilizando el método del capítulo 11. Sin embargo, el modelo estructural es más flexible y puede aplicarse a estudios con muestras de tamaños desiguales.

Los cálculos realizados a través del modelo estructural usualmente están resumidos en una *tabla del análisis de varianza*, con una columna para la fuente de variación (intergrupal, intragrupal y total), las sumas de desviaciones cuadráticos (SC), los grados de libertad (gl), las estimaciones de la varianza poblacional (CM, que es igual a SC/gl) y F (que es igual a CM_{Entre}/CM_{Dentro}).

Los supuestos son los mismos que los de cualquier análisis de varianza, aunque el análisis con grupos de tamaños desiguales es un poco más sensible a los incumplimientos de los supuestos.

Un análisis de varianza es seguido generalmente de comparaciones múltiples, planificadas o *post hoc*, las cuales analizan las diferencias entre pares o subgrupos específicos de medias. Dichas comparaciones tienen que protegerse contra la posibilidad de obtener algunos resultados significativos sólo por casualidad, debido a que pueden realizarse una gran cantidad de comparaciones.

La proporción de varianza explicada (R^2), también denominada eta cuadrado (η^2), es una medida del tamaño del efecto del análisis de varianza. Es SC_{Entre} dividida por SC_{Total}.

Algunos expertos recomiendan que en lugar de utilizar un análisis de varianza para realizar comparaciones difusas y generales entre varias medias, los investigadores deberían planificar previamente la realización de comparaciones planificadas específicas, apuntadas directamente a las cuestiones teóricas.

Términos clave

- Tabla del análisis de varianza.
- Procedimiento Bonferroni.
- Eta cuadrado (η^2).
- Comparaciones múltiples.
- Comparaciones planificadas.
- Comparaciones *post hoc*.
- Proporción de varianza explicada (R^2).
- Modelo estructural.
- Suma de desviaciones cuadráticas (SC_{Entre}, SC_{Dentro}, SC_{Total}).

Ejercicios

Los ejercicios implican la realización de cálculos (con la ayuda de una calculadora). La mayoría de los problemas estadísticos reales se resuelven por computadora, pero aunque exista la posibilidad de utilizarla, es conveniente realizar estos ejercicios manualmente para incorporar el método de trabajo.

Para adquirir práctica en la utilización de una computadora, para resolver problemas estadísticos, se puede utilizar la sección de computación de cada capítulo, publicada en la *Guía de estudio y libro de tareas de computación para el alumno [Student's Study Guide and Computer Workbook]* que acompaña este libro.

Todos los datos de esta sección son ficticios (a menos que se especifique lo contrario).

Las respuestas a los ejercicios de la serie I se encuentran al final del libro.

SERIE I

1. Los datos mostrados a continuación son los mismos que aparecen en el ejercicio 2 de la serie I del capítulo11. Resuelva el mismo problema utilizando el método del modelo estructural y compare su respuesta con la respuesta lograda en el capítulo 11 (utilice el nivel 0,01). Asegúrese de mostrar sus cálculos y de incluir una tabla del análisis de varianza.

Grupo 1	Grupo 2	Grupo 3
8	6	4
8	6	4
7	5	3
9	7	5

2. Calcule un análisis de varianza para los siguientes datos (al nivel de significación del 1%). Asegúrese de mostrar sus cálculos y de incluir una tabla del análisis de varianza.

Grupo 1	Grupo 2	Grupo 3	Grupo 4
71	82	68	78
67	82	70	76
	82		

3. Para cada una de los siguientes conjuntos de datos, calcule a) las medias de cada grupo, b) un análisis de varianza utilizando el método del modelo estructural (al nivel de significación del 5%) y c) R^2. (Al realizar el paso b, asegúrese de mostrar todos sus cálculos y de incluir una tabla del análisis de varianza).

	Grupo 1	Grupo 2	Grupo 3
(i)	3	0	1
	4	1	2
	5	2	3
	Grupo 1	**Grupo 2**	**Grupo 3**
(ii)	3	0	1
	5	0	3
	1	2	2

4. Un investigador está interesado en los niveles de autoestima de profesores de tres materias diferentes. Los niveles de autoestima de los cuatro profesores de lengua analizados fueron 2, 2, 3 y 5. Los niveles de autoestima de los tres profesores de matemática analizados fueron 6, 4 y 5. Los niveles de autoestima de los cinco profesores de ciencias sociales analizados fueron 9, 6, 7, 10 y 13. ¿Sostienen los resultados una diferencia en los niveles medios de autoestima de los tres distintos tipos de profesores (al nivel 0,05)?

a) Realice el análisis de varianza. b) Calcule R^2. c) Explique su respuesta a alguien que comprende los conceptos de media, varianza y varianza poblacional estimada (incluyendo las nociones de muestra, población y grados de libertad), pero que no sabe nada más sobre estadística.

5. Un estudio comparaba la efectividad de los programas de prevención del abuso de drogas. En toda Norteamérica existen cuatro programas que utilizan el método A, en el cual sus puntuaciones atribuidas generales son 13, 8, 10 y 9; tres programas que utilizan el método B, en el cual sus puntuaciones atribuidas son 5, 7 y 6, y otros tres programas que utilizan el método C, en el cual sus puntuaciones atribuidas son 4, 5 y 2. Sobre la base de esas muestras, ¿deberíamos concluir que los programas que utilizan diferentes métodos tienen diferentes grados de efectividad? Utilice el nivel 0,05. Escriba un informe a una comisión del gobierno explicando sus conclusiones. El informe debería escribirse de forma tal que lo comprendan funcionarios que tal vez nunca hayan asistido a un curso sobre estadística.

6. Van Lange et al. (1997) realizaron un estudio en el que los participantes tomaban parte en una tarea de juegos estándar. En la tarea de juegos, el participante realiza una serie de decisiones en cuanto a otorgarse puntos a sí mismo o a otra persona. Utilizando los resultados de esta tarea, se puede clasificar a cada participante según su "orientación en cuanto a valores sociales" como "pro-social" (tiende a ser cooperativo y favorecer resultados igualitarios para sí mismo y para otros), "individualista" (busca obtener la mayor cantidad posible para él mismo sin preocuparse por el resultado logrado por otros) o "competitivo" (se preocupa porque su resultado sea mejor al de los demás). Una de las hipótesis de Van Lange et al. establecía que las personas "pro-sociales" tendrían más hermanos que los integrantes de los otros dos grupos. Sus resultados mostraron un efecto significativo general, "$F(2, 535) = 4{,}82$, $p < 0{,}01$". (*p*. 739) Luego, lo informan de la siguiente forma:

> Coherentemente con la hipótesis que relaciona hermanos y carácter pro-social de la persona [...] la cantidad de hermanos es mayor en el caso de las personas pro-sociales ($M = 2{,}03$, $SD = 1{,}56$) que en caso de los individualistas ($M = 1{,}63$, $SD = 1{,}00$) y los competitivos ($M = 1{,}71$, $SD = 1{,}35$). Comparaciones planificadas realizadas posteriormente revelaron un contraste significativo entre pro-sociales versus individualistas y competitivos, $F(1, 535) = 9{,}14$, $p < 0{,}005$. Las diferencias entre individualistas y competitivos no fueron significativas. (pp. 739–740)

Explique el significado de todo lo anterior a alguien que nunca ha tomado un curso de estadística.

SERIE II

1. El ejercicio 3 de la serie II del capítulo 11 era un análisis de varianza que investigaba si los individuos que trabajaban en diferentes áreas de una empresa tenían diferentes actitudes hacia la misma. Los resultados, en cuanto a actitudes positivas para las tres personas analizadas del área de ingeniería, fueron 10, 12 y 11; para los tres del área de comercialización, fueron, 6, 6 y 8; para los tres de área de contaduría, 7, 4 y 4 y para los tres del área de producción, 14, 16 y 13. Resuelva el mismo problema utilizando el método del modelo estructural y compare su respuesta con la respuesta del capítulo 11. (Asegúrese de mostrar todos sus cálculos y de incluir una tabla completa del análisis de varianza).

2. Calcule un análisis de varianza para los siguientes datos (al 5% de nivel de significación). Asegúrese de mostrar sus cálculos y de incluir una tabla completa del análisis de varianza.

Grupo 1	Grupo 2	Grupo 3	Grupo 4	Grupo 5
1	3	2	2	4
7	11	12	8	10
		6		8

3. Calcule un análisis de varianza para cada uno de los siguientes conjuntos de datos, utilizando el método del modelo estructural para cada una (a un nivel del 5% de significación). Además, calcule R^2. Asegúrese de mostrar sus cálculos y de incluir una tabla completa del análisis de varianza para cada una.

	Grupo 1	Grupo 2	Grupo 3
(i)	0	0	4
	2	2	6
		0	
		2	

	Grupo 1	Grupo 2	Grupo 3
(ii)	0	0	4
	2	2	6
			5
			6

4. Un investigador dedicado al tema del descanso comparó el efecto de tres tipos de alteraciones del sueño (ser despertado en distintos intervalos) en la agilidad mental al día siguiente. Originalmente, había 12 participantes en la investigación que fueron asignados al azar a una de las tres condiciones (4 por condición). Sin embargo, uno de los participantes del programa I de alteración no cumplió las instrucciones, y la información proveniente de ese participante no pudo utilizarse en el análisis, lo cual dio como resultado muestras de tamaños desiguales. Los resultados de la medida de agilidad mental fueron las siguientes: programa I de alteración: 120, 140, 140; programa II de alteración: 130, 150, 120, 140; programa III de alteración: 100, 90, 110, 120. ¿Sostienen los resultados un efecto diferente en la agilidad mental producido por los tres tipos de programas de alteración (al nivel 0,05)? a) Realice el análisis de varianza. b) Calcule R^2. c) Explique su respuesta a alguien que comprende los conceptos de media, varianza y varianza estimada de población (incluidas las nociones de muestra, población y grados de libertad) pero que no sabe nada más sobre estadística.

5. El ejercicio 5 de la serie II del capítulo 10 era una prueba t para medias independientes, correspondiente a un estudio acerca de los efectos del color para calmar la angustia. Comparaba los valores en las pruebas de angustia realizadas por individuos en papel amarillo pastel o papel verde chillón. Los valores observados de los cinco participantes que realizaron la prueba impresa en papel amarillo fueron 17, 19, 28, 21 y 18. Los valores de los cuatro participantes que completaron la prueba en el papel verde fueron 20, 26, 17 y 24. Calcule un análisis de varianza con estos datos. (Está utilizando el ANOVA para una situación con sólo dos grupos). Si saca la raíz cuadrada de la razón F, debería obtener lo

mismo (teniendo en cuenta el error de redondeo) que el valor t que calculó utilizando la prueba t para medias independientes (veremos esta relación entre la prueba t y el análisis de varianza en el capítulo 16).

6. Cialdini y sus colegas (1997) pidieron a sus participantes que indiquen hasta qué grado ayudarían a una persona en problemas. Por ejemplo, en una pregunta se les pedía que indicaran cuánta ayuda le darían a una persona que acababa de ser desalojada de su departamento. Las posibles respuestas iban desde "no hacer nada", pasando por opciones intermedias tales como "llevar en auto a la persona a buscar un nuevo departamento", hasta opciones extremas tales como "invitar a la persona a vivir con uno indefinidamente". Se asignaba a los participantes a una de las cuatro condiciones según quién fuera la persona necesitada: a) casi un extraño, b) un conocido, c) un buen amigo, d) un familiar cercano. Cialdini et al. informaron:

> Realizamos un ANOVA probando nuestras expectativas generales en cuanto a que una relación más cercana aumentaría la voluntad de ayudar. Ese análisis produjo un efecto altamente significativo que sostenía nuestra hipótesis, $F(3, 82) = 33{,}28$, $p < 0{,}001$. La [tabla 12-10] presenta las medias de ayuda relacionadas con cada uno de los niveles de cercanía de la relación.

Explique los resultados a una persona que nunca ha tomado un curso sobre estadística. (Concéntrese sólo en la línea superior de la tabla, en los resultados en cuanto a la ayuda del estudio 1).

Tabla 12-10.
Medias de registros en cuanto a la ayuda, la preocupación empática y la entereza como una función del nivel de cercanía de la relación y de la situación de necesidad.

	Nivel de cercanía de la relación			
Situación de necesidad	Casi extraño	Conocido	Buen amigo	Familiar cercano
Estudio 1: Desalojo				
Ayuda	$1{,}20_a$	$4{,}13_b$	$6{,}63_c$	$6{,}89_c$
Preocupación empática	$3{,}04_a$	$4{,}36_b$	$4{,}21_b$	$4{,}50_b$
Entereza	$1{,}52_a$	$3{,}16_b$	$4{,}52_c$	$4{,}57_c$
n	22	22	20	22
Estudio 2: Niños huérfanos				
Ayuda	$4{,}13_a$	$6{,}11_a$	$7{,}96_b$	$9{,}01_b$
Preocupación empática	$4{,}42_a$	$5{,}52_b$	$5{,}85_b$	$5{,}82_b$
Entereza	$1{,}90_a$	$3{,}18_b$	$5{,}24_c$	$4{,}66_b$
n	15	17	17	19
Estudio 3: Llamada telefónica				
Ayuda	$0{,}80_a$	$0{,}98_{a'b}$	$1{,}54_b$	$1{,}55_b$
Preocupación empática	$2{,}87_a$	$3{,}49_{a'b}$	$4{,}55_b$	$4{,}66_b$
Entereza	$2{,}17_a$	$3{,}16_a$	$4{,}43_b$	$4{,}66_b$
n	33	18	20	19
Desalojo				
Ayuda	$1{,}77_a$	$3{,}63_b$	$5{,}88_c$	$6{,}95_c$
Preocupación empática	$3{,}56_a$	$4{,}34_{a'b}$	$4{,}90_{b'c}$	$5{,}66_c$
Entereza	$2{,}16_a$	$3{,}56_b$	$5{,}00_c$	5,66d
n	27	19	20	16
Niños huérfanos				
Ayuda	$4{,}15_a$	$5{,}36_a$	$8{,}23_b$	$8{,}83_b$
Preocupación empática	$4{,}53_a$	$4{,}51_a$	$5{,}41_{a'b}$	$6{,}21_b$
Entereza	$2{,}40_a$	$3{,}02_a$	$4{,}48_b$	$4{,}80_b$
n	20	23	19	20

Nota: Dentro de cada fila, las medias que tienen el mismo subíndice no son significativamente diferentes según la prueba de Tukey.

Fuente: Cialdini, R. B., Brown, S. L., Lewis, B. P., Luce, C., & Neuberg, S. L. (1997), tab. 1. "Reinterpretación de la relación empatía-altruismo: cuando uno en uno es igual a entereza". *Revista de Psicología Social y de Personalidad [Journal of Personality and Social Psychology]*, 73, 481–494. Copyright, 1997, por la Asociación Americana de Psicología. Reimpreso con autorización.

Apéndice I del capítulo: fórmulas de cálculo optativas para las sumas de cuadrados en un análisis de varianza de un criterio

Las siguientes son fórmulas de cálculo para las sumas de los cuadrados:

$$SC_{\text{Total}} = \Sigma X^2 - \frac{(\Sigma X)^2}{N} \tag{12-6}$$

$$SC_{\text{Entre}} = \frac{(\Sigma X_1)^2}{n_1} + \frac{(\Sigma X_2)^2}{n_2} + \cdots + \frac{(\Sigma X_{\text{Último}})^2}{n_{\text{Último}}} - \frac{(\Sigma X)^2}{N} \tag{12-7}$$

$$SC_{\text{Dentro}} = SC_{\text{Total}} - SC_{\text{Entre}} \tag{12-8}$$

X_1, X_2... $X_{\text{Último}}$ son los valores observados en cada grupo, y N_1, N_2,..., $n_{\text{Último}}$ son la cantidad de observaciones de cada grupo.

La tabla 12-11 muestra los cálculos del último ejemplo completo del capítulo, con la aplicación de estas fórmulas. Compare estos cálculos con los indicados en la tabla 12-4, en la que se aplican fórmulas de definición.

Tabla 12-11.
Análisis de varianza de valores de ansiedad basado aproximadamente en Clark et al. (1997), en el que se aplican fórmulas de cálculo para las sumas de las desviaciones cuadráticas. (Datos ficticios).

	No pacientes		Pacientes con pánico		Pacientes con angustia generalizada		Pacientes con fobia social	
	X	X^2	X	X^2	X	X^2	X	X^2
	7	49	11	121	10	100	11	121
	8	64	10	100	12	144	11	121
	10	100	12	144			11	121
	7	49						
Σ:	32	262	33	365	22	244	33	363

$\Sigma X = 32 + 33 + 22 + 33 = 120$

$\Sigma X^2 = 262 + 365 + 244 + 363 = 1{,}234$

$$SC_{\text{Total}} = \Sigma X^2 - \frac{(\Sigma X)^2}{N} = 1.234 - \frac{120^2}{12} = 1.234 - \frac{14.400}{12} = 1.234 - 1.200 = 34$$

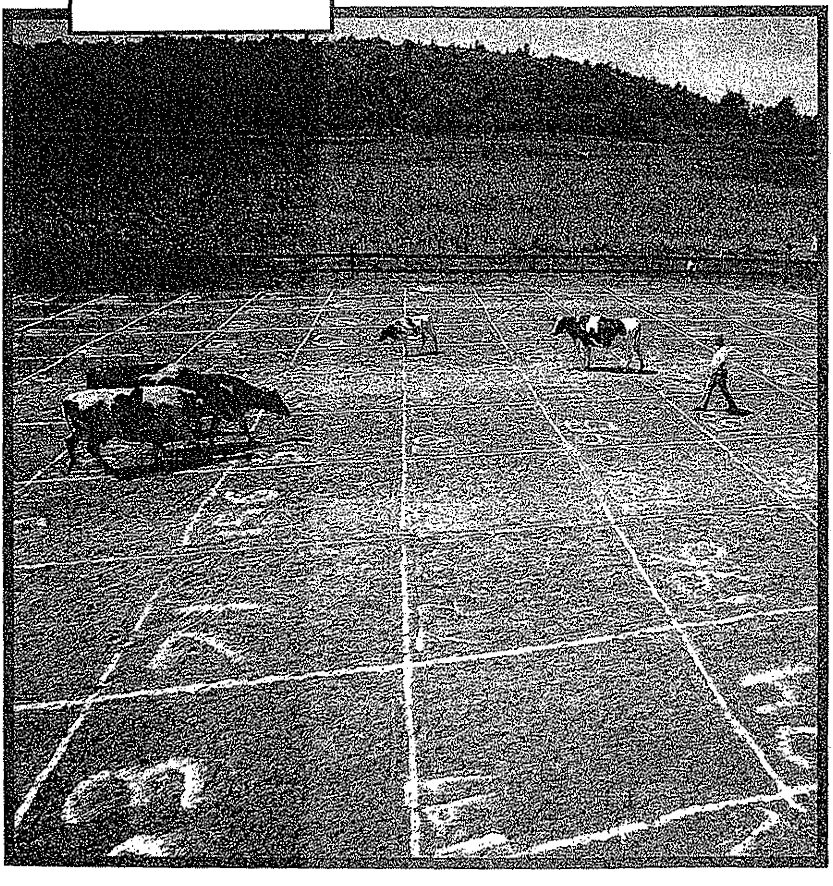
$$SC_{\text{Entre}} = \frac{(\Sigma X_1)^2}{n_1} + \frac{(\Sigma X_2)^2}{n_2} + \ldots + \frac{(\Sigma X_{\text{Última}})^2}{n_{\text{Última}}} - \frac{(\Sigma X)^2}{N} = \frac{32^2}{4} + \frac{33^2}{3} + \frac{22^2}{2} + \frac{33^2}{3} - \frac{120^2}{12}$$

$$= \frac{1.024}{4} + \frac{1.089}{3} + \frac{484}{2} + \frac{1.089}{3} - \frac{14{,}440}{12}$$

$$= 256 + 363 + 242 + 363 - 1{,}200 = 24$$

$$SC_{\text{Dentro}} = SC_{\text{Total}} - SC_{\text{Entre}} = 34 - 24 = 10$$

13 Análisis de Varianza Factorial



Descripción del capítulo

- Lógica básica de los diseños factoriales y los efectos interactivos.
- Lógica básica del análisis de varianza de dos criterios de clasificación.
- Potencia y tamaño del efecto del análisis factorial de varianza.
- Extensiones y casos especiales del análisis factorial de varianza.
- Controversias, limitaciones y desarrollos recientes.
- Los resultados del análisis factorial de varianza según se describen en las publicaciones científicas.
- Resumen.
- Términos clave.
- Ejercicios.
- Apéndice I del capítulo: fórmulas de cálculo optativas para el análisis de varianza de dos criterios.
- Apéndice II del capítulo: análisis de varianza de un criterio con medidas repetidas.

En este capítulo presentamos el análisis de varianza factorial, una extensión de los procedimientos aprendidos en los capítulos 11 y 12. El análisis de varianza factorial proporciona un enfoque altamente flexible y eficiente para analizar resultados de cierto tipo de experimentos complejos que son ampliamente utilizados en psicología.

Comenzaremos el capítulo analizando con detenimiento la naturaleza de estos complejos diseños factoriales de investigación; luego expondremos brevemente el razonamiento y los procedimientos de cálculo de un análisis de varianza factorial. En este capítulo hemos invertido la presentación del material por una buena razón. La lógica y la terminología de los diseños experimentales probablemente sean nuevos para el alumno, mientras que el razonamiento y los procedimientos de cálculo involucrados en la realización de un análisis de varianza factorial son una extensión bastante directa de lo aprendido en el capítulo 12.

LÓGICA BÁSICA DE LOS DISEÑOS FACTORIALES Y DE LOS EFECTOS INTERACTIVOS

Presentaremos el análisis de varianza factorial a través de un ejemplo. Lambert y sus colegas (1997) estaban interesados en la forma en que los estereotipos afectan las evaluaciones que hacemos de los demás. Por ejemplo, con frecuencia las personas aplican estereotipos relacionados con la edad y el sexo para evaluar si alguien tendrá éxito en un determinado empleo. Lambert et al. estaban especialmente interesados en la forma en que la influencia de los estereotipos se ve afectada por a) la conciencia de que el estereotipo no es adecuado para una determinada circunstancia y b) nuestro estado de ánimo. Creían que las personas se ven afectadas en menor medida por los estereotipos cuando éstos no son apropiados y, en especial, que no se ven afectadas por ellos cuando se sienten tristes.

Por lo tanto, Lambert et al., realizaron el siguiente experimento. Se pidió a los participantes que actuaran como entrevistadores laborales. Su tarea era "realizar una evaluación preliminar acerca de la conveniencia de un individuo para determinado empleo" (p. 1010), puesto que en todos los casos era el de un asistente de vuelo. Después se entregó a los participantes un currículum de un solicitante que incluía la foto de una mujer muy atractiva y, sobre la base de esa información, se preguntaba a los participantes cuáles eran las posibilidades de que la contrataran, utilizando una escala del 0 (para nada) al 10 (extremadamente). El experimento utilizó el estereotipo del atractivo físico, que incluye la tendencia a pensar que las personas atractivas son especialmente competentes.

Los investigadores lograron que la mitad de los participantes se sintieran tristes antes de leer el currículum, supuestamente como parte de otro experimento. A esos participantes se les pidió que pensaran en "un episodio de sus vidas que los haya hecho sentir muy tristes y, que incluso hoy en día, los sigue poniendo tristes cuando lo recuerdan" (p. 1004) Lo que acabamos de describir era la condición de tristeza. La otra mitad de los participantes, a quienes no se les dio ninguna instrucción en particular, formaron la condición neutra.

La segunda influencia de interés para los investigadores era el hecho de que el estereotipo fuera el apropiado. Se entregó a los participantes una descripción de un buen asistente de vuelo, que variaba en cuanto al grado de importancia del atractivo físico para el puesto. Para la mitad de los participantes de cada uno de los grupos, la descripción subrayaba la capacidad "de resolver y analizar problemas de forma racional y analítica" (p. 1010); esta era la condición del estereotipo inapropiado. Para los otros participantes, la descripción subrayaba la satisfacción de los pasajeros y la forma en que la apariencia contribuía a ello; esta era la condición del estereotipo apropiado.

En resumen, había dos manipulaciones experimentales: tristeza en contraposición a neutro, y la descripción de puesto, adecuada al estereotipo, en contraposición con la inadecuada.

Lambert y sus colegas podrían haber realizado dos estudios: uno comparando participantes tristes con neutros, y otro comparando participantes que recibieron descripciones de puesto adecuadas al estereotipo con los de las descripciones no adecuadas al estereotipo. Pero en lugar de eso, decidieron analizar en un sólo estudio los efectos de ambas circunstancias, es decir, el estado de ánimo y la calidad del estereotipo. Analizaron cuatro grupos de participantes (*véase* tabla 13-1): a) aquellos en condición de tristeza y con el estereotipo apropiado, b) aquellos en condición de tristeza y con el estereotipo inapropiado, c) aquellos en condición neutra y con el estereotipo apropiado y d) aquellos en condición neutra y con el estereotipo inapropiado.

Tabla 13-1.
Diseño factorial utilizado por Lambert et al. (1997).

		Estado de ánimo	
		Triste	*Neutro*
Estereotipo	*Apropiado*	a	c
	Inapropiado	b	d

Definición del diseño factorial de investigación

El estudio de Lambert et al. (1997) es un ejemplo de un estudio **con diseño factorial de investigación**, en el que se analiza de una sola vez el efecto de dos o más variables formando grupos con cada combinación de dichas variables. En el ejemplo que hemos presentado, existen dos niveles de estado de ánimo (triste y neutro) y dos niveles de calidad del estereotipo (apropiado e inapropiado), que permiten cuatro combinaciones posibles. Lambert et al. utilizaron todas ellas en su estudio.

Un diseño factorial de investigación presenta una importante ventaja con respecto a la realización de estudios de cada variable por separado: la eficiencia. Con un diseño factorial podemos analizar ambas variables de una sola vez, sin necesidad de convocar el doble de participantes. En el ejemplo presentado, Lambert et al. pudieron utilizar un sólo grupo de participantes para analizar los efectos del estado de ánimo y de la calidad del estereotipo.

Efectos interactivos

Existe una ventaja aún más importante del diseño factorial de investigación. Este diseño brinda la posibilidad de analizar los efectos de la combinación de dos o más variables. En el ejemplo que estamos analizando, el estado de ánimo y la calidad del estereotipo podrían afectar la contratación en un modo simple y aditivo. Lo que queremos decir es que las influencias combinadas podrían ser la suma de las influencias separadas; por lo tanto, si aumenta una de esas influencias, y también la otra, entonces el efecto general, que es la suma total de los dos efectos individuales, también será mayor. Por ejemplo, supongamos que sentirse triste predispone a contratar a alguien y, similarmente, el hecho de que el estereotipo sea apropiado predispone a contratar a una persona. Si estos dos efectos son simplemente aditivos, entonces los participantes del grupo que se siente triste y que recibió el estereotipo apropiado serán los más predispuestos a contratar a la persona; los participantes del grupo neutro, que recibieron el estereotipo inapropiado, serán los que tengan menos predisposición a contratar a la persona, y aquellos en las otras dos condiciones tendrían una predisposición intermedia para contratar a la persona en cuestión.

También podría suceder que una variable tuviera cierto efecto y la otra no. O que tal vez ninguna variable tuviera ningún efecto. En la situación aditiva, o en la que sólo una variable o ninguna tienen efectos, observar a las variables en combinación no agrega ninguna información interesante.

Sin embargo, también es posible que la combinación de las dos variables cambie el resultado. De hecho, Lambert et al. predijeron que el efecto del estereotipo inapropiado sería especialmente fuerte en la condición de tristeza. Esta predicción se basaba en la noción de que cuando nos sentimos tristes, estamos más predispuestos a rever nuestras reacciones iniciales, irreflexivas, basadas en un estereotipo.

El anterior es sólo un ejemplo del modo en que diferentes condiciones podrían combinarse en formas que no esperaríamos al conocer sólo el efecto de cada factor separadamente. Veamos otro ejemplo. Supongamos que sentirse triste disminuye la posibilidad de contratar a alguien cuando el estereotipo es inapropiado, pero que aumenta la probabilidad de contratar cuando el estereotipo es apropiado (es decir, que la tristeza hace que las personas presten más atención a lo apropiado de la situación). Incluso, otra posibilidad sería que el único grupo que tenga altas probabilidades de contratar a la persona sea aquel en el que el estereotipo es apropiado y el estado de ánimo neutro, es decir, estando triste, se clasifica con bajo nivel a todos en general, pero con un estado de ánimo neutro, se presta atención a lo adecuado del estereotipo y se clasifica con un nivel alto a aquél cuyo

Tabla 13-2.
Posibilidad media de contratación en el estudio de Lambert et al. (1997).

		Estado de ánimo	
Estereotipo		*Triste*	*Neutro*
	Apropiado	7,73	5,80
	Inapropiado	5,83	6,75

atractivo físico es adecuado. Existen también otras posibilidades (sería interesante que el alumno pensara algunas y razonara su significado en relación con los temas que estamos estudiando).

Todas aquellas situaciones en las cuales la **combinación** de variables tiene un efecto especial son ejemplos de lo que se denomina **efecto interactivo.** Un efecto interactivo ocurre cuando el efecto de una variable depende del nivel de la otra variable. En el estudio de Lambert et al., existía un efecto interactivo. Si observamos la tabla 13-2 veremos que el resultado fue que los participantes del grupo apropiado-triste presentaban las mayores posibilidades de contratar al solicitante; el grupo inapropiado-neutro le seguía en orden de posibilidad, y los otros dos grupos eran los que tenían menos posibilidades (estando casi al mismo nivel). Analicemos la parte del resultado en el que el grupo triste-inapropiado presenta menos posibilidades de contratar que el grupo neutro-inapropiado. Esta parte del resultado sostiene la teoría del investigador que establece que, cuando uno se siente triste, puede contrarrestar sus estereotipos. (¿Qué sucede con el hecho de que el grupo apropiado-triste sea el que presenta más posibilidades de contratar? Los investigadores reconocieron que este resultado era "inesperado y difícil de explicar" (p. 1011).

Supongamos que los investigadores habían analizado la calidad del estereotipo y el estado de ánimo en dos estudios separados. Habrían llegado a la conclusión de que cada factor tenía sólo un leve efecto. El promedio de posibilidad de contratación, siendo el estereotipo apropiado, es 6,77 (es decir, el promedio de 5,80 y 7,73 es igual a 6,77), y en los casos en los que el estereotipo es inapropiado, el promedio es 6,29. La posibilidad promedio de contratación, en el caso de los que se sienten tristes, es de 6,78, contra 6,28 de aquellos en la condición neutra. Por lo tanto, siguiendo el método de los dos estudios independientes, los resultados importantes se hubieran pasado totalmente por alto. Los resultados más importantes estaban relacionados con la combinación de los dos factores. ¿Ya se empieza a percibir la importancia de descubrir los efectos de la interacción?

Terminología

El estudio de Lambert et al. se analizaría con lo que se denomina un **análisis de varianza de dos criterios de clasificación** (el que se aplica a un **diseño factorial de investigación de dos factores**). Por el contrario, las situaciones que analizamos anteriormente en los capítulos 11 y 12 (como el estudio acerca del estilo de vinculación o el experimento con antecedentes delictivos) eran ejemplos de estudios analizados utilizando un **análisis de varianza de un criterio**. Se dice que estos estudios tienen un criterio único de clasificación porque analizan el efecto de una sola variable (como por ejemplo, el efecto del estilo de vinculación de una persona o de la información acerca de los antecedentes delictivos del acusado).

Algunos estudios investigan el efecto de tres o más variables juntas. Por ejemplo, Lambert et al. también querían estar seguros de que sus resultados no estaban afectados por el sexo. Por lo

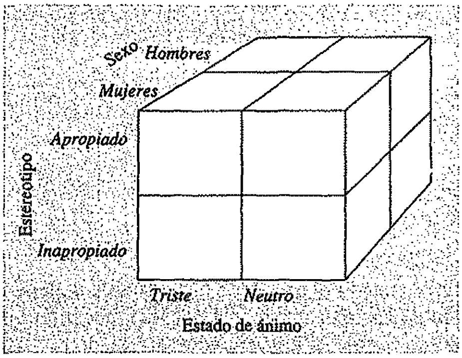


Figura 13-1.
Diseño factorial de tres factores utilizado en el estudio de Lambert et al. (1997).

tanto, en otro análisis dividieron cada uno de sus cuatro grupos en dos subgrupos: mujeres y hombres. A través de esa división se crearon ocho combinaciones: mujeres tristes con el estereotipo apropiado, hombres tristes con el estereotipo apropiado, mujeres tristes con el estereotipo inapropiado, y así sucesivamente. La figura 13-1 representa gráficamente la serie completa de agrupaciones. El análisis que describimos estudiaba la influencia de tres variables al mismo tiempo. Se requieren tres dimensiones para diagramar este tipo de estudios, por eso se lo denomina **diseño factorial de tres factores.** (El resultado indicó que no se producían efectos significativos relacionados con el sexo, ni generales ni en interacción con el estado de ánimo, ni con la calidad del estereotipo o la combinación de los dos). Es posible realizar diseños factoriales de cuatro factores o más, aunque no son sencillos de diagramar. Sin embargo, la mayoría de las investigaciones psicológicas se limitan a diseños factoriales de dos factores y, ocasionalmente, de tres.

En un análisis de dos criterios, cada variable o "criterio de clasificación" (cada dimensión en el diagrama) es un posible efecto principal. Si el resultado de una variable, haciendo un promedio a través de las categorías de la otra u otras variables, es significativo, se trata de un **efecto principal.** Lo anterior es completamente diferente de un efecto interactivo, que se basa en la combinación de variables. En el estudio básico de Lambert et al. en dos sentidos, existía la posibilidad de dos efectos principales y de un efecto interactivo. Los dos efectos principales posibles son la calidad del estereotipo y el estado de ánimo, y el efecto interactivo posible es el de la combinación de la calidad del estereotipo con el estado de ánimo. En un análisis de varianza de dos criterios, siempre se prueban dos posibles efectos principales y una posible interacción.

Cada combinación de grupos en un diseño factorial se denomina **casilla.** La media de los valores observados de cada agrupación se denomina **media de la casilla.** Por ejemplo, en el estudio de Lambert et al. existen cuatro casillas, por ende, existen cuatro medias de casillas, una para cada combinación de los niveles de calidad del estereotipo y del estado de ánimo. Es decir, una casilla se refiere al estereotipo apropiado y al estado de ánimo triste (como lo muestra la tabla 13-2, su media es 7,73); otra casilla se refiere al estereotipo inapropiado y al estado de ánimo triste (5, 83); otra casilla se refiere al estereotipo apropiado y el estado de ánimo neutro (5,80), y otra casilla se refiere al estereotipo inapropiado y el estado de ánimo neutro (6, 75).

Las medias según una sola variable se denominan **medias marginales.** Por ejemplo, en el estudio de Lambert et al. hay cuatro medias marginales, una media correspondiente a todos los participantes que trabajan con el estereotipo apropiado (como vimos anteriormente, 6,77), una para todos los participantes que trabajan con el estereotipo inapropiado (6,29), una para todos los participantes que se sienten tristes (6,78) y una para todos los participantes con estado de ánimo neutro (6,28). (Las medias que acabamos de mencionar no aparecen en las tablas porque estábamos interesados principalmente en la interacción).

Para observar el efecto principal debemos concentramos en las medias marginales. Para observar el efecto interactivo, debemos concentrarnos en el patrón de medias de las casillas individuales.

Reconocimiento e interpretación de los efectos interactivos

Es muy importante comprender los efectos interactivos, ya que en muchos experimentos constituyen el punto principal de la investigación.

Como ya hemos visto, un efecto interactivo ocurre cuando el efecto de una variable depende del nivel de otra variable. Los resultados del estudio de Lambert et al. (tabla 13-2) muestran un efecto interactivo, ya que el efecto de la calidad del estereotipo es diferente con un estado de ánimo triste que con un estado de ánimo neutro.

Un efecto interactivo puede explicarse de tres modos: con palabras, con números o con un gráfico. Podemos describir un efecto interactivo con palabras diciendo que ocurre cuando el efecto de una variable depende del nivel de otra variable. En nuestro ejemplo de Lambert et al. podemos decir que el efecto de la calidad del estereotipo depende del nivel del estado de ánimo (también podemos decir que el efecto del estado de ánimo depende del nivel de la calidad del estereotipo. Los efectos interactivos son completamente simétricos, es decir que podemos describirlos desde el punto de vista de cualquiera de las variables).

Podemos observar un efecto interactivo numéricamente analizando el patrón de las medias de las casillas. Si existe un efecto interactivo, las diferencias de las medias de las casillas de una fila no serán iguales a las diferencias de las medias de las casillas de la otra fila. Analicemos el ejemplo de Lambert et al. En la fila del estereotipo apropiado hay una diferencia positiva entre las medias de las casillas; los participantes que se sentían tristes clasificaron las posibilidades de contratación (7,33) mucho más altas que los participantes con estado de ánimo neutro (5,80). En consecuencia, hay una diferencia positiva de 1,93 (es decir, 7,73 – 5,80 = 1,93). Sin embargo, si observamos la fila del estereotipo inapropiado, aquellos que se sentían tristes calificaron la posibilidad de su probabilidad de contratación (5,83) a **menor nivel** que aquellos con estado de ánimo neutro (6,75). La diferencia entre estado de ánimo triste y neutro, para los participantes con el estereotipo inapropiado, es de –0,92.

La tabla 13-3 indica las medias marginales y de casilla correspondientes a seis posibles resultados de un estudio factorial ficticio de dos criterios, que buscaba la relación de la edad y la educación con el ingreso. La edad se divide en dos niveles (menores, entre 25 y 29; en contraposición con mayores, aquellos que tienen entre 30 y 34) y la educación en otros dos niveles (secundaria en contraposición con universitaria). Los resultados ficticios que presentamos en este caso están exagerados con el fin de dejar en claro cuándo hay efectos interactivos y cuándo efectos principales. En la realidad, con frecuencia se encuentran pequeñas diferencias de medias en la dirección de un efecto interactivo o uno principal, que no son lo suficientemente grandes como para ser estadísticamente significativas.

Tabla 13-3.
Posibles medias de resultados de un estudio acerca de la relación de la edad y la educación con el ingreso. (Datos ficticios, miles de dólares).

Resultado A

	Educación secundaria	*Educación universitaria*	*Total*
Menores	20	20	20
Mayores	20	30	25
Total	20	25	

Resultado B

	Educación secundaria	*Educación universitaria*	*Total*
Menores	30	20	25
Mayores	20	30	25
Total	25	25	

Resultado C

	Educación secundaria	*Educación universitaria*	*Total*
Menores	10	30	20
Mayores	20	40	30
Total	15	35	

Resultado D

	Educación secundaria	*Educación universitaria*	*Total*
Menores	10	10	10
Mayores	60	60	60
Total	35	35	

Resultado E

	Educación secundaria	*Educación universitaria*	*Total*
Menores	20	30	25
Mayores	20	40	30
Total	20	35	

Resultado F

	Educación secundaria	*Educación universitaria*	*Total*
Menores	20	30	25
Mayores	30	50	40
Total	25	40	

En el resultado A, existe una interacción. Observamos que en la fila "Menores" no existe diferencia de ingresos por educación; pero en la fila "Mayores", la media de la casilla correspondiente a la educación universitaria es mucho mayor que la media de la casilla correspondiente a educación secundaria. Una manera de expresar verbalmente lo anterior sería la siguiente: "La educación no está relacionada con el ingreso en el caso del grupo 'Menores', pero en el caso del grupo 'Mayores', las personas con educación universitaria ganan mucho más que aquellas con menor nivel de educación".

El resultado ficticio B también refleja una interacción. En la fila "Menores", la media de ingreso correspondiente a la educación secundaria es mayor a la media de ingreso correspondiente a la educación universitaria; sin embargo, en la fila "Mayores" la media de ingreso correspondiente a la "educación secundaria" es menor. Expresado verbalmente, este patrón indica que entre las personas de menor edad, aquellos con sólo una educación secundaria ganan más dinero (tal vez porque ingresaron al empleo con anterioridad o porque las clases de empleos que desempeñan comienzan con un nivel superior); sin embargo, entre las personas de mayor edad, aquellos con una educación universitaria ganan más dinero.

El resultado ficticio C **no** refleja un efecto interactivo. En la fila "Menores", la media de educación secundaria es 20 puntos menor que la media de educación universitaria. Lo mismo ocurre en la fila "Mayores". Expresado en palabras, lo anterior significa que, ya sean menores o mayores, las personas con educación universitaria ganan $20.000 más.

El resultado ficticio D tampoco refleja interacción, ya que en ninguna de las filas existe diferencia alguna. Independientemente de la educación, las personas mayores ganan $50.000 más.

El resultado ficticio E refleja una interacción. En la fila "Menores", la media correspondiente a la educación universitaria es 10 puntos mayor; pero en la fila "Mayores", la media correspondiente a la educación universitaria es 20 puntos mayor. Por lo tanto, si bien entre las personas menores, los que tienen educación universitaria ganan un poco más, entre las personas mayores, los que tienen educación universitaria ganan mucho más.

Finalmente, el resultado F también refleja un efecto interactivo. Existe una diferencia menor en la fila "Menores" que en la fila "Mayores". Al igual que el resultado E, este patrón indica que, en el caso de las personas menores, aquellas que tienen educación universitaria ganan un poco más; pero entre las personas mayores, aquellas con educación universitaria ganan mucho más.[1]

La tabla 13-4 indica los posibles resultados de otro estudio ficticio. En este experimento factorial, las dos variables manipuladas experimentalmente son el grado de dificultad de la tarea (fácil en contraposición con difícil) y el nivel de excitación psicológica (baja, moderada o alta). La excitación, en este estudio, se refiere al nivel de ansiedad del participante con respecto a la importancia de realizar bien la tarea. La variable que se está midiendo es el nivel de desempeño del participante en una serie de tareas aritméticas. La interpretación de las posibles interacciones es la siguiente:

Resultado A: no hay interacción. Las medias de las casillas en la fila "Fácil" no difieren entre sí, y las medias de las casillas en la fila "Difícil" tampoco. Sí existe un efecto principal: la dificultad de la tarea afecta el desempeño; la excitación no.

Resultado B: no hay interacción. Las medias de las casillas en la fila "Fácil" aumentan de a 10, de bajo a moderado y de moderado a alto. Lo mismo ocurre con las medias de las casillas en la fila "Difícil". Nuevamente existe sólo un efecto principal: la excitación afecta el desempeño; la dificultad de la tarea, no.

Resultado C: no hay interacción. Las medias de las casillas en la fila "Fácil" aumentan de a 10, de bajo a moderado y de moderado a alto; lo mismo sucede con las medias de las casillas en la fila "Difícil". En este ejemplo, existen dos efectos principales: la excitación afecta el desempeño y la dificultad de la tarea también.

Resultado D: existe interacción. El patrón de las medias de las casillas en la fila "Fácil" muestra un aumento de a 10, de bajo a moderado, y otro aumento de a 10, de moderado a alto. El patrón que describimos anteriormente no es el mismo que el de las medias de las casillas en la fila "Difícil", donde nuevamente el aumento es de a 10 de bajo a moderado, pero de 40 de moderado a alto. Por lo tanto, en todos los casos, el desempeño con tareas fáciles o difíciles tiende a mejorar con el aumento de la excitación. Sin embargo, el impacto entre excitación alta y moderada es mucho mayor para tareas difíciles que para tareas fáciles.

Resultado E: existe interacción. El patrón de las medias de las casillas en la fila "Fácil" muestra un aumento de a 10 y luego una disminución de a 10. Este patrón es bastante diferente al de la fila "Difícil", donde observamos una disminución de a 10 y luego un aumento de a 10. En el caso de las tareas fáciles, el mejor desempeño se produce en la condición de excitación moderada; en cambio, para las tareas difíciles, el peor desempeño se produce en la condición de excitación moderada.

Resultado F: existe interacción. En la fila "Fácil", las medias de las casillas aumentan a medida que avanzamos, mientras que en la fila "Difícil" disminuyen. En el caso de las tareas fáciles, a mayor excitación, mejor es el desempeño; en el caso de las tareas difíciles, la excitación interfiere con el desempeño. (El resultado *F* es el más cercano a un descubrimiento psicológico bien fundamentado, el cual se conoce como la Ley de Yerkes-Dodson).

[1] Sobre la base de las estadísticas realizadas en 1990 por la Secretaría de Educación de los Estados Unidos, la situación actual en ese país es muy cercana al resultado *F*, aunque no tan extrema. En ambas franjas de edad, las personas con educación universitaria ganan más que aquellas que tienen sólo educación secundaria, pero la diferencia es un poco mayor en la franja de personas mayores. Sin embargo, es importante recordar que el hecho de que una persona reciba o no educación universitaria también está relacionado con la clase social de sus padres y con otros factores que pueden afectar el ingreso más de lo que lo hace la educación.

Tabla 13-4.
Algunos resultados posibles de un experimento acerca del efecto del grado de dificultad y la excitación sobre el desempeño. (Datos ficticios).

	Resultado A				Resultado B				Resultado C			
	Excitación				*Excitación*				*Excitación*			
Tarea	*Baja*	*Moderada*	*Alta*	*Total*	*Baja*	*Moderada*	*Alta*	*Total*	*Baja*	*Moderada*	*Alta*	*Total*
Fácil	10	10	10	10	10	20	30	20	10	20	30	20
Difícil	20	20	20	20	10	20	30	20	20	30	40	30
Total	15	15	15		10	20	30		15	25	35	

	Resultado D				Resultado E				Resultado F			
Fácil	10	20	30	20	10	20	10	13,3	10	20	30	20
Difícil	10	20	60	30	20	10	20	16,7	30	20	10	20
Total	10	20	45		15	15	15		20	20	20	

Identificación gráfica de los efectos interactivos

Otra forma común de interpretar los efectos interactivos es mediante la representación gráfica del patrón de las medias de las casillas. Por lo general, la representación gráfica se realiza a través de un gráfico de barras.[2] La figura 13-2 es una reproducción de la publicación del estudio de Lambert et al. Los gráficos de las figuras 13-3 y 13-4 muestran los gráficos de los resultados ficticios que acabamos de analizar (los que aparecen en las tablas 13-3 y 13-4, respectivamente).

Acerca de los gráficos que mencionamos arriba cabe destacar lo siguiente: siempre que existe una interacción, el patrón de las barras de una sección del gráfico será diferente del patrón de la otra sección del gráfico. Así, en la figura 13-2, el patrón correspondiente a inapropiado-triste está un escalón más abajo, pero el patrón correspondiente a apropiado-triste está un escalón más arriba. El hecho de que las barras tengan diferentes patrones es precisamente una forma gráfica de indicar que el patrón de diferencias entre las medias de las casillas de una fila y de la otra no es el mismo.

Analicemos la figura 13-3. Observemos primero los resultados C y D. En los resultados C, las series de barras que representan los grupos "Menores" y "Mayores" tienen el mismo patrón, ambas aumentan de a 20. En el resultado D, ambas son parejas. En los resultados C y D, las barras que representan al grupo "Menores" y las que representan al grupo "Mayores" tienen el mismo patrón. El C y el D eran los ejemplos en donde no había interacción. En todos los otros resultados, en los que sí había interacción, los patrones de las barras no son paralelos. Por ejemplo, en el resultado A, las dos barras que representan al grupo "Menores" están parejas, mientras que las ba-

[2] La utilización de gráficos de barras para representar las medias de las casillas de un análisis de varianza, cuando existe un efecto interactivo, se ha convertido en el método estándar en los últimos años. Anteriormente, era más común utilizar gráficos de líneas. Veremos este tema y ejemplos referidos al mismo en la sección donde tratamos la descripción del análisis factorial de varianza en las publicaciones científicas.

rras que representan al grupo "Mayores" indican un aumento. En los resultados B, las barras que representan al grupo "Menores" muestran una disminución, en el caso de la educación universitaria, con respecto a la secundaria, mientras que las barras que representan el grupo "Mayores" muestran un aumento en el caso de la educación universitaria con respecto a la secundaria. En los resultados E y F, tanto las barras que representan al grupo "Menores" como al "Mayores" muestran un aumento, pero las barras que representan al grupo "Menores" muestran un aumento menor que las barras que representan al grupo "Mayores".

Analicemos la figura 13-4. Los resultados A, B y C muestran la ausencia de interacción dentro de cada resultado, y los patrones de las barras que representan la excitación baja, moderada y alta son iguales. En el resultado D existe interacción, que se refleja en la figura de la siguiente manera: las barras dentro del nivel bajo son parejas al igual que dentro del nivel moderado; pero entre las barras que representan el nivel de excitación alto, hay un aumento en la condición de tarea difícil con respecto a la tarea fácil. La interacción del resultado E se ve en los aumentos en los niveles bajo y alto de excitación y, por otro lado, en la disminución en el nivel moderado de excitación. La interacción del resultado F se refleja en el hecho de que existe un aumento entre las barras del nivel bajo, mientras que las barras del nivel moderado son parejas, y una disminución en la barra que representa la condición difícil.

La figura 13-5 muestra una alternativa diferente a la de la figura 13-4 en cuanto a la forma de representar gráficamente los resultados de la tabla 13-4. En este caso, hemos agrupado las barras correspondientes a las condiciones "difícil" y "fácil". Las barras que representan la tarea fácil con excitación baja, moderada y alta están ubicadas una al lado de la otra, y las barras que representan la tarea difícil, con excitación baja, moderada y alta, se encuentran también una al lado de la otra. Esta alternativa de agrupación es completamente equivalente en significado y produce exactamente las mismas conclusiones. Por ejemplo, en el resultado A las tres barras de tarea difícil son parejas, al igual que las tres barras de tarea fácil. En el resultado C, donde tampoco hay interacción, las tres barras de tarea fácil aumentan con el mismo patrón que las tres barras de tarea difícil. Sin embargo, analicemos el resultado D, donde sí existe una interacción. El patrón de las

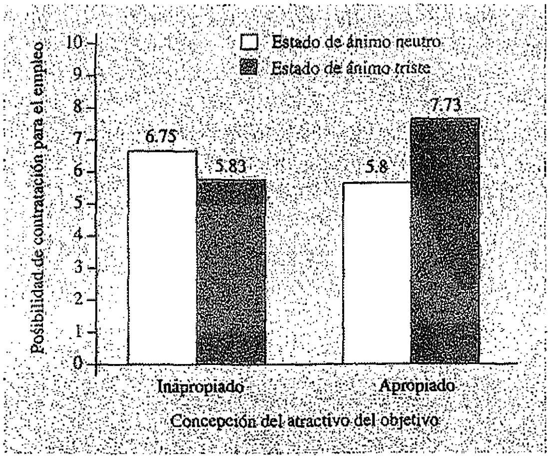


Figura 13-2.
Calificaciones con respecto a la candidata físicamente atractiva para el empleo, como una función de la concepción del atractivo físico (inapropiado en contraposición con apropiado) y del estado de ánimo manipulado (triste en contraposición con neutro). Experimento 3. Los números mayores indican una mayor posibilidad de contratación del objetivo para el empleo [**Fuente:** Lambert, A. J.Khan, S. R., LickelB. A. & Fricke, K. (1997), fig. 1. "El estado de ánimo y la correlación con estereotipos positivos en contraposición con los negativos". *Revista científica de Psicología Social y de Personalidad [Journal of Personality and Social Psychology]* 72, 1002-1016. Copyright 1997, por la Asociación Americana de Psicología. Reimpreso con Autorización]

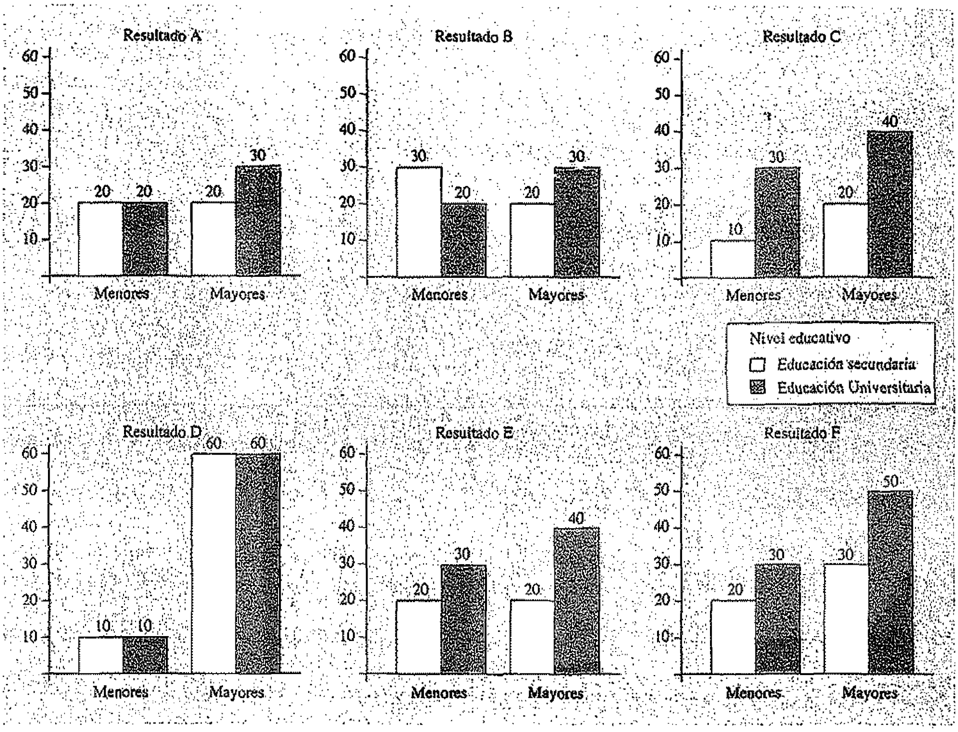


Figura 13-3.
Gráficos de los resultados ficticios de la tabla 13-3.

barras que representan la tarea fácil es diferente del patrón de las barras que representan la tarea difícil. Existe un escalón mayor entre la excitación moderada a alta en las barras de la tarea difícil que en las de tarea fácil.

También podemos identificar efectos principales a partir de estos gráficos. En la figura 13-3, se mostraría un efecto principal de la edad si las barras que representan al grupo de menor edad fueran todas más altas o más bajas que las barras que representan al grupo de mayor edad. Por ejemplo, en el resultado C, las barras que representan al grupo de mayor edad son claramente más altas que las barras del grupo de menor edad. ¿Qué sucede con el efecto principal de las barras que no están agrupadas, como son, en este caso, la educación universitaria en contraposición con la secundaria? En el caso de las barras no agrupadas, debemos observar si el patrón general aumenta o disminuye. Por ejemplo, en el resultado C también existe un efecto principal de la educación, porque el patrón general de las barras aumenta en cuanto a la educación universitaria con respecto a la educación secundaria, y lo hace tanto para el grupo de menor edad como para el de mayor edad. El resultado D muestra un efecto principal de la edad (las barras del grupo de mayor edad son más altas que las barras del grupo de menor edad). Sin embargo, el resultado D no mues-

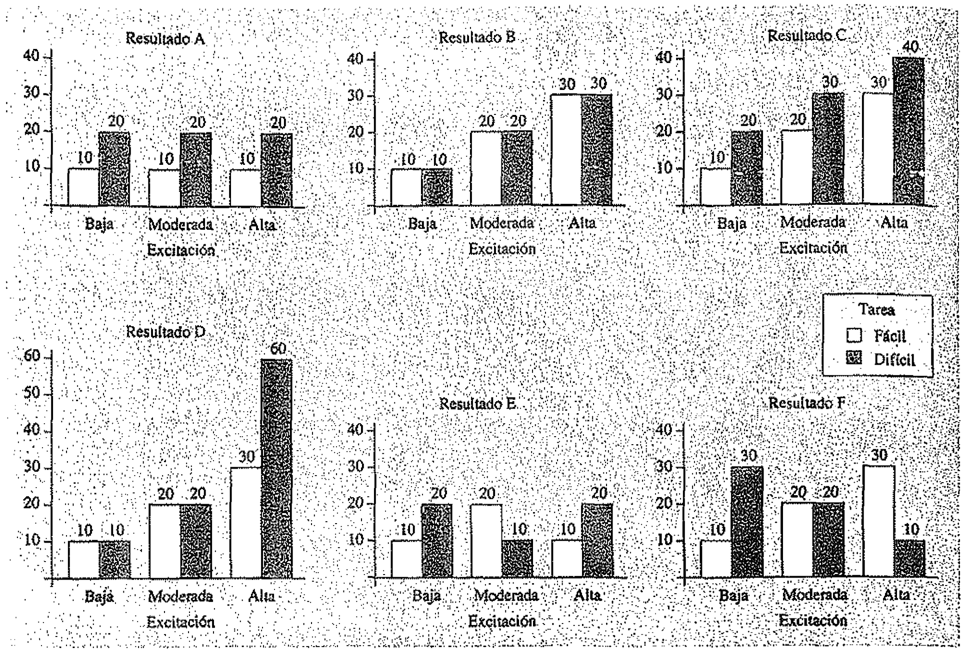


Figura 13-4.
Gráficos de los resultados ficticios de la tabla 13-4.

tra un efecto principal de la educación; el patrón es parejo tanto para las barras que representan al grupo de mayor edad como para las que representan al grupo de menor edad. El resultado A en la figura 13-4 muestra un efecto principal del grado de dificultad de la tarea, pero no del nivel de excitación, ya que las alturas promedio de las barras son las mismas para la excitación baja, moderada y alta; mientras que dentro de cada nivel de excitación, las barras aumentan de fácil a difícil.

Relación entre efectos interactivos y principales

Es posible que cualquier combinación de efectos principales e interactivos sea significativa. Por ejemplo, puede ocurrir que todas sean significativas, como lo muestra el patrón del resultado F en la tabla 13-3, según el cual observamos que los alumnos mayores ganan más (efecto principal de edad), los alumnos universitarios ganan más (efecto principal del nivel de educación), y el nivel hasta el cual los alumnos universitarios ganan más depende de la edad (efecto interactivo).

De modo similar, en el resultado D de la tabla 13-4, vemos que, en líneas generales, las personas se desempeñan mejor en tareas difíciles (efecto principal relacionado con la dificultad de la tarea) y con mayores niveles de excitación (efecto principal relacionado con el nivel de excitación), pero el efecto producido por la dificultad de la tarea sólo se observa en relación con altos niveles de excitación (interacción). (Cabe destacar, de todos modos, que el efecto principal producido por la dificultad de la tarea, es decir, el mayor promedio logrado al realizar tareas más di-

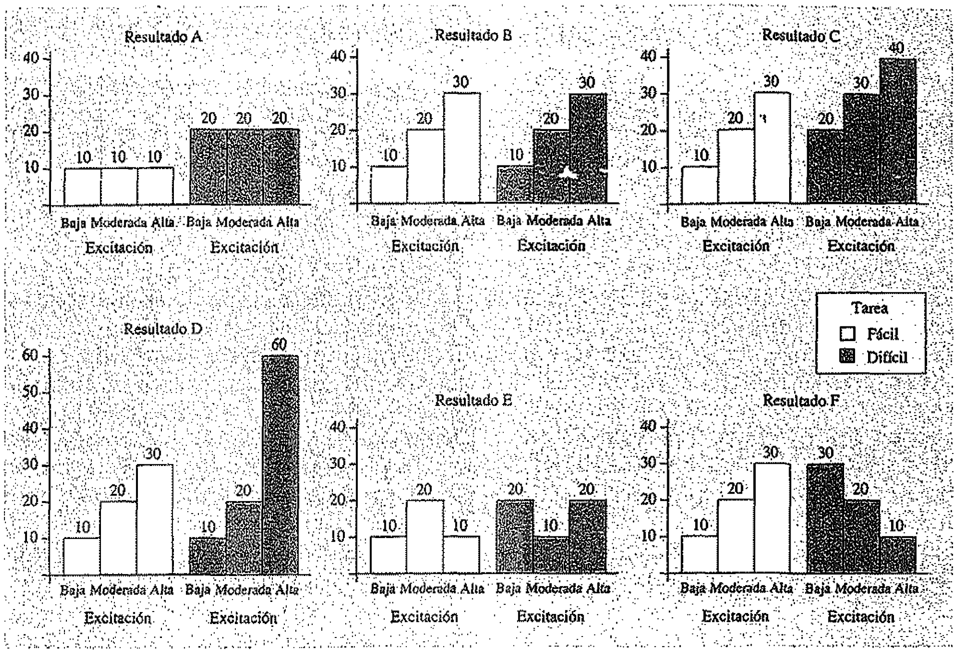


Figura 13-5.
Gráficos alternativos (en relación con la figura 13-4) de los resultados ficticios que aparecen en la tabla 13-4.

fíciles, se debe en su totalidad al alto nivel de excitación. Más adelante seguiremos tratando este tipo de situaciones).

También puede existir un efecto interactivo sin efectos principales. El resultado B de la tabla 13-3 es un ejemplo de lo anterior. El nivel promedio de ingreso es el mismo para "Menores" y "Mayores" (no existe efecto principal de la edad), y es el mismo para personas con educación universitaria o secundaria (no existe efecto principal del nivel de educación). De manera similar, en el resultado F de la tabla 13-4, el desempeño promedio es el mismo para los niveles bajo, moderado y alto de excitación (no existe efecto principal del nivel de excitación), y es el mismo para tareas fáciles y difíciles (no existe efecto principal del nivel de dificultad de la tarea). Sin embargo, en ambos ejemplos existen claras interacciones.

El ejemplo de Lambert et al. (1997) que analizamos anteriormente es, en realidad, un ejemplo de interacción sin efectos principales (*véase* tabla 13-2 ó figura 13-2). Es verdad, que en líneas generales, los participantes que se sentían tristes calificaron sus posibilidades de contratar de forma más alta que los participantes con estado de ánimo neutro. Sin embargo, la diferencia no fue lo suficientemente importante como para resultar significativa desde el punto de vista estadístico. De modo similar, la diferencia entre las condiciones en las que el estereotipo era apropiado o inapropiado no fue lo suficientemente importante como para ser significativa. Es decir, que en ese estudio sólo el efecto interactivo era significativo. (En la próxima sección principal veremos cómo se calcula realmente si un efecto es lo suficientemente importante como para ser significativo).

También puede ocurrir que exista un efecto principal significativo junto con una interacción, o sólo un efecto principal significativo, o bien que no existan ni efectos principales ni interactivos que sean significativos. Sería interesante observar cuántas de esas posibilidades se pueden identificar en las dos series de resultados ficticios de las tablas 13-3 y 13-4.

Cuando no existe interacción, el efecto principal tiene un significado directo. Sin embargo, cuando existe una interacción junto con un efecto principal, debemos ser cuidadosos al sacar conclusiones acerca del efecto principal. Analicemos el resultado D del ejemplo relacionado con la excitación y la dificultad de la tarea (tabla 13-4). Suponiendo que las diferencias son lo suficientemente importantes como para ser significativas, existen dos efectos principales y una interacción. Pero como observamos anteriormente, el efecto principal de la dificultad de la tarea se debe completamente a la casilla en la que se combinan el alto nivel de excitación y la tarea difícil. Sería engañoso realizar cualquier afirmación acerca de la comparación de tareas difíciles con tareas fáciles en general, sin tener en cuenta que *el efecto realmente depende del nivel de excitación*.

A veces, el efecto principal se mantiene claramente por encima de cualquier interacción. Analicemos nuevamente el resultado D del ejemplo acerca de excitación y dificultad de la tarea. En ese resultado, el efecto principal de la excitación se mantiene por encima de la interacción. Tanto en el caso de las tareas fáciles como de las difíciles ocurre que el bajo nivel de excitación produce el menor nivel de desempeño, la excitación moderada produce el segundo desempeño en la escala, y el nivel alto de excitación produce el mejor desempeño. (Aun así, existe una interacción porque el grado en el cual un alto nivel de excitación produce mejor desempeño que la excitación moderada es mayor para las tareas difíciles que para las tareas fáciles).

LÓGICA BÁSICA DEL ANÁLISIS DE VARIANZA DE DOS CRITERIOS

El procedimiento estadístico para analizar los resultados de un experimento factorial en dos sentidos se denomina análisis de varianza de dos criterios. La lógica básica es la misma que la que aprendimos en los capítulos 11 y 12. En cualquier análisis de varianza se calcula una razón F; y esa razón F compara la estimación de la varianza poblacional, basada en la variación **entre** las medias de los grupos de interés, con la estimación de la varianza poblacional basada en la variación **dentro** de los grupos.

Las tres razones F de un análisis de varianza de dos criterios

En un análisis de varianza de dos criterios, existen tres razones F: uno para el efecto principal de la columna, uno para el efecto principal de la fila y uno para el efecto interactivo. El numerador de cada una de estas razones F será una estimación intergrupal de la varianza poblacional, basada en los grupos que se comparan al analizar determinado efecto interactivo o principal. La estimación intragrupal de varianza es la misma para las tres razones F: siempre será el promedio de las estimaciones de varianza poblacional calculadas a partir de las observaciones internas de cada una de las casillas.

Lógica de la determinación de las razones F para los efectos principales de columnas y de filas

Una manera de comprender cómo se realiza el análisis de los efectos principales es la siguiente. Analicemos el efecto principal de columnas. Calculemos la siguiente razón F: el numerador es una estimación intergrupal de la varianza basada en la variación entre las medias marginales de columnas. El denominador es una estimación intragrupal de la varianza basada en el promedio de las estimaciones de la varianza a partir de cada una de las casillas. Pensemos en el ejemplo del estudio de Lambert et al. (1997). La razón F correspondiente al estado de ánimo (la variable de las columnas según como hemos diagramado el cuadro) se calcula de la siguiente forma. El numerador, la estimación intergrupal de la varianza, se basa en la diferencia entre la media marginal del estado de ánimo triste y la media marginal del estado de ánimo neutro. El denominador, la estimación intragrupal de la varianza, se basa en el promedio de las estimaciones de la varianza poblacional calculadas a partir del interior de las cuatro casillas.

El procedimiento para el efecto principal de las filas mantiene la misma lógica. Se calcula utilizando una estimación intergrupal de la varianza basada en la diferencia entre las dos medias marginales de las filas. (En el estudio de Lambert et al. las medias marginales de las filas son la media de todos los participantes con estereotipo apropiado y la media de todos los participantes con estereotipo inapropiado)

La figura 13-6a muestra la "estimación intergrupal columna" de la varianza, basada en la diferencia entre la media de las observaciones de la primera columna (área sombreada) y la media de las observaciones de la segunda columna (área sin sombrear). La figura13-6b muestra la "estimación intergrupal fila", basada en la diferencia entre la media de las observaciones de la fila superior (área sombreada) y la media de las observaciones de la fila inferior (área no sombreada). Finalmente, la figura 13-6c muestra la "estimación intragrupal de la varianza" (utilizada para todas las razones F), basada en la variación entre las observaciones de cada una de las casillas.

Lógica de la determinación de la razón F de los efectos interactivos

La lógica de la razón F de un efecto interactivo es un poco más compleja. Se trata también de una razón entre una estimación intergrupal y una estimación intragrupal. La estimación intragrupal es el promedio usual de las estimaciones de la varianza poblacional, calculado a partir de todas las casillas individuales. Lo que resulta más complejo es el cálculo de la estimación intergrupal de la varianza a partir del efecto interactivo.

Una forma consiste en considerar al efecto interactivo como una descripción de las posibles combinaciones restantes después de analizar los efectos principales de las columnas y las filas. Es decir, en un diseño 2 x 2, los efectos principales han agrupado a las cuatro casillas en filas y co-

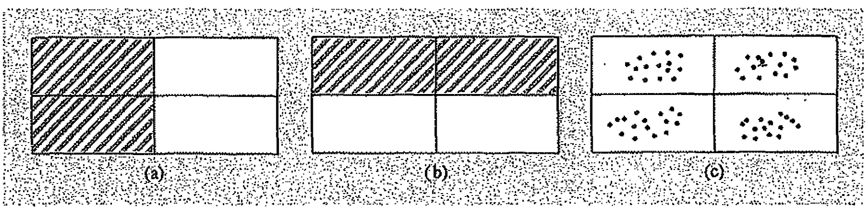


Figura 13-6.
Diagrama de ayuda para comprender un análisis de varianza factorial 2 x 2: (a) estimación intergrupal columna de la varianza basada en la diferencia entre la media de los participantes en la primera (sombreada) y la segunda (no sombreada) columna; (b) estimación intergrupal fila, basada en la diferencia entre la media de los participantes de la fila superior (sombreada) e inferior (no sombreada), y (c) estimación intragrupal de varianza basada en la variación entre las observaciones de cada casilla.

Cuadro 13-1.
Influencias de la personalidad y las circunstancias en el comportamiento: un efecto interactivo.

En el cuadro 12-1 vimos que el análisis de varianza simula la forma en que los investigadores en el campo de la psicología planifican la investigación, como también la forma en que todos pensamos. Conociendo ese paralelismo –sea que los investigadores realicen la comparación conscientemente o no– es probable que con frecuencia utilicen el modelo bien definido del análisis de varianza como guía de su propia lógica. Y lo hacen no sólo cuando analizan información o cuando diseñan una investigación; probablemente, también utilizan el análisis de varianza como metáfora cuando teorizan. Estudiar estadística es, en cierto sentido, un entrenamiento en cuanto a la manera de ver el mundo.

Un claro ejemplo del modo en que la estadística influye en la forma de pensar de los psicólogos acerca de su objeto de estudio, y no sólo sobre la información recopilada, es el estudio de la personalidad. En los años 60, el campo de la personalidad cambió para siempre a través del trabajo de Walter Mischel (1968). Mischel parecía haber demostrado que, como regla general, la **circunstancia** (el semáforo poniéndose en rojo, por ejemplo, o una persona bien vestida pidiendo ayuda) era un mejor elemento de predicción de cómo va a actuar una persona que cualquier otro rasgo de la personalidad (por ejemplo, que una persona sea cautelosa o altruista por naturaleza). Acosados, los teóricos de la personalidad –que eran de formación psicodinámica– lucharon por defenderse dentro de las reglas de juego que Mischel había establecido: ¿Cuánto de la varianza del comportamiento podía realmente predecirse a través de medidas de la personalidad? Es decir, que los teóricos de la personalidad se vieron obligados a pensar estadísticamente.

Uno de los resultados de este desafío ha sido lo que se dio en llamar "interaccionismo" (p. ej. Endler & Magnusson, 1976). El interaccionismo representa la idea de que el comportamiento se predice mejor a través de la interacción entre la persona y la circunstancia. De inmediato adivinamos cuál fue el método estadístico que mayor influencia tuvo en este campo (lo estamos analizando en este capítulo).

Por ejemplo, de acuerdo con este modelo, ni la ansiedad como rasgo de personalidad, ni la circunstancia de rendir el SAT, es tan buen elemento de predicción del estado de ansiedad como saber que una persona con determinada tendencia a la ansiedad percibe el hecho de rendir el SAT como una circunstancia angustiante. El acento está puesto en el hecho de que el comportamiento es alterado constantemente por la disposición interna del individuo en interacción con su percepción de las cambiantes circunstancias.

Sigamos a un hombre ansioso a través de algunas situaciones. Puede sentirse aún más o menos ansioso a medida que pasa de la circunstancia del examen a una playa de estacionamiento oscura y vacía, según la interacción que se produzca entre su rasgo de ansiedad y su percepción de esa nueva circunstancia. Lo mismo sucederá cuando vaya en camino hacia su casa en la autopista, abra el portón del garaje y entre en una casa vacía.

De acuerdo con el interaccionismo, la persona no es un componente pasivo sino "un agente activo intencional en el proceso de interacción" (Endler & Magnusson, 1976 p. 968). Lo importante de la circunstancia es, nuevamente, su significado para la persona.

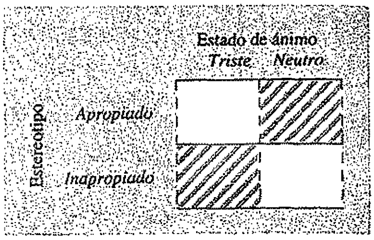
Los interaccionistas admiten que este tipo de modelo estadístico es aún demasiado mecánico y lineal. En el mundo real, existe un intercambio constante entre la circunstancia y la persona, algo más parecido a una transacción que a una interacción. No obstante, sostienen que para probar estos modelos se necesitan herramientas estadísticas más complicadas. Y están llegando; por ejemplo, Kenny (1995) predice que dentro de 10 años el análisis basado en el historial de sucesos y el modelo de niveles múltiples serán herramientas estándar para el análisis de información. Por eso, a medida que los expertos en estadística produzcan metodologías más complicadas, los teóricos de la personalidad las adoptarán, no sólo como herramientas para el análisis de información sino también como modelos de la influencia mutua entre el interior de las persona y su mundo exterior.

Esa misma influencia de la estadística en la teoría se repite en el área de la cognición, la percepción y el aprendizaje (Gigerenzer & Murray (1987), entre muchas otras áreas de la psicología. En cierto sentido, podríamos decir que los pioneros de la estadística ahora están determinando no sólo la complejidad posible de la investigación psicológica sino también la profundidad de la propia formulación teórica. Están adueñándose de los canales a través de los cuales fluye el patrón real de pensamiento de los psicólogos y, por lo tanto, al menos en la actualidad, están dando forma y dirigiendo gran parte de nuestra comprensión de la psicología.

lumnas, pero también es posible agrupar las casillas de un modo diferente. La figura 13-7, basada en el estudio de Lambert et al., muestra una posible organización restante de las cuatro casillas en dos agrupaciones mayores: a) una agrupación de dos casillas formada por la casilla superior izquierda (estereotipo apropiado y estado de ánimo triste) junto con la casilla inferior derecha (estereotipo inapropiado y estado de ánimo neutro), y b) otra agrupación de dos casillas formada por la casilla inferior izquierda (estereotipo inapropiado y estado de ánimo triste) y la casilla superior derecha (estereotipo apropiado y estado de ánimo neutro). La estimación intergrupal de la varianza, según el efecto interactivo, puede entonces ser determinada a partir de la variación entre las medias de estas dos agrupaciones.

Con un diseño 2 x 2, existe sólo una combinación de pares de casillas que ya no fueron tenidas en cuenta por las organizaciones en columnas y en filas, el patrón de agrupación representado por el ejemplo en la figura 13-7. Pero con diseños mayores de dos criterios, como por ejemplo un diseño 2 x 3, existe más de una forma de combinar las agrupaciones, y todas deben tenerse en cuenta. Por eso, calcular la estimación intergrupal de la varianza a partir de un efecto interactivo, cuando tratamos con situaciones distintas del diseño 2 x 2, puede ser bastante complicado. Afortunadamente, sucede que calcular la estimación intergrupal de la varianza, según el efecto de la interacción, es mucho más directo desde el punto de vista del modelo estructural que aprendimos en el capítulo 12.

Figura 13-7.
Interacción como comparación de la media de las observaciones de las casillas sombreadas (estado de ánimo neutro, estereotipo apropiado y estado de ánimo triste, estereotipo inapropiado) con la media de las observaciones en las casillas no sombreadas (estado de ánimo triste, estereotipo apropiado y estado de ánimo neutro, estereotipo inapropiado) del estudio de Lambert et al. (1997).



Modelo estructural para el análisis de varianza de dos criterios

Desde la perspectiva del modelo estructural, el desvío general de cada observación con respecto a la gran media puede dividirse en varios componentes. En un análisis de dos criterios, existen cuatro componentes de ese desvío general (*véase* también la figura 13-8):

1. El desvío de las observaciones con respecto a la media de su casilla (que se utiliza en la estimación intragrupal de la varianza poblacional).
2. El desvío de la media de la fila de la observación con respecto a la gran media (que se utiliza en la estimación intergrupal de la varianza poblacional a partir del efecto principal de la variable diagramada a lo largo de las filas).

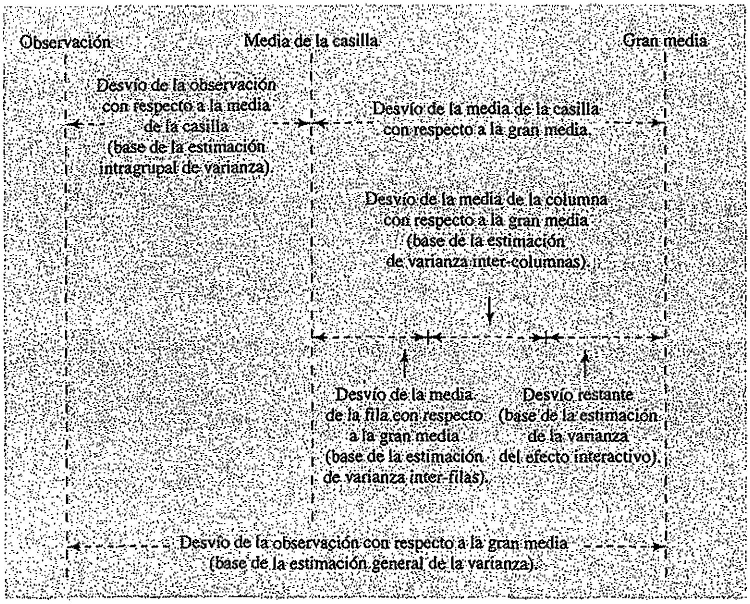


Figura 13-8.
Cálculo del desvío de cada observación con respecto a la gran media.

3. El desvío de la media de la columna de la observación con respecto a la gran media (que se utiliza en la estimación intergrupal basada en el efecto principal de la variable diagramada a lo largo de las columnas).

4. Un desvío restante, que queda después de restar los otros tres desvíos al desvío general con respecto a la gran media (es el que se utiliza en la estimación intergrupal del efecto interactivo).

Es recomendable tomarse un momento para estudiar la figura 13-8, ya que es la mejor manera de comprender y recordar lo que estamos explicando.

Cálculos del análisis de varianza de dos criterios

Al realizar un análisis de varianza utilizando el modelo estructural, las razones F se obtienen a) calculando todos los desvíos de cada tipo, b) elevándolos al cuadrado, c) sumando los desvíos cuadráticos de cada tipo para obtener las sumas de los desvíos cuadráticos, d) dividiendo cada suma de desvíos cuadráticos por los grados de libertad correspondientes para obtener las estimaciones de la varianza y e) dividiendo las distintas estimaciones de la varianza intergrupales por las estimaciones de la varianza intragrupales.

Las sumas de los cuadrados se calculan con las siguientes fórmulas:

$$SC_{\text{Filas}} = \Sigma(M_{\text{Filas}} - GM)^2 \tag{13-1}$$

$$SC_{\text{Columnas}} = \Sigma(M_{\text{Columnas}} - GM)^2 \tag{13-2}$$

$$SC_{\text{Interacción}} = \Sigma[(X - GM) - (X - M) - (M_{\text{Filas}} - GM) - (M_{\text{Columnas}} - GM)]^2 \tag{13-3}$$

$$SC_{\text{Dentro}} = \Sigma(X - M)^2 \tag{13-4}$$

$$SC_{\text{Total}} = \Sigma(X - GM)^2 \tag{13-5}$$

En las fórmulas anteriores, SC_{Filas}, SC_{Columnas}, $SC_{\text{Interacción}}$ y SC_{Dentro} son las sumas de los cuadrados de las filas, de las columnas, de las interacciones y de las intragrupales de las casillas. El signo suma (Σ) indica que se deben sumar todas las observaciones que corresponden (no sólo todas las filas o columnas o casillas). GM es la gran media; X es la observación; M_{Fila} y M_{Columna} son las medias de la fila o la columna de una observación, y M es la media de la casilla de una observación.

Como es usual, las diferentes sumas individuales de cuadrados forman la suma total de cuadrados. (Este dato se puede utilizar para controlar los cálculos aritméticos). La fórmula es la siguiente,

$$SC_{\text{Total}} = SC_{\text{Filas}} + SC_{\text{Columnas}} + SC_{\text{Interacción}} + SC_{\text{Dentro}} \tag{13-6}$$

Las fórmulas de las estimaciones de la varianza poblacional son, por lo general, las sumas de los cuadrados divididas por los grados de libertad:

$$S^2_{\text{Filas}} \text{ ó } CM_{\text{Filas}} = \frac{SC_{\text{Filas}}}{gl_{\text{Filas}}} \tag{13-7}$$

$$S^2_{\text{Columnas}} \text{ ó } CM_{\text{Columnas}} = \frac{SC_{\text{Columnas}}}{gl_{\text{Columnas}}} \tag{13-8}$$

$$S^2_{\text{Interacción}} \text{ ó } CM_{\text{Interacción}} = \frac{SC_{\text{Interacción}}}{gl_{\text{Interacción}}} \tag{13-9}$$

$$S^2_{\text{Dentro}} \text{ ó } CM_{\text{Dentro}} = \frac{SC_{\text{Dentro}}}{gl_{\text{Dentro}}} \tag{13-10}$$

En las fórmulas anteriores, SC_{Filas} ó CM_{Filas} es la estimación de la varianza poblacional calculada a partir de las filas; S^2_{Columnas} ó CM_{Columnas} es la estimación de varianza poblacional calculada a partir de las columnas; $S^2_{\text{interacción}}$ ó $CM_{\text{interacción}}$ es la estimación de varianza poblacional calculada a partir de la interacción; S^2_{Dentro} ó CM_{Dentro} es la estimación intragrupal de varianza poblacional.

Las razones F se calculan dividiendo las estimaciones de la varianza poblacional, calculadas a partir de los diferentes efectos, por la estimación intragrupal de la varianza poblacional:

$$F_{\text{Filas}} = \frac{S^2_{\text{Filas}}}{S^2_{\text{Dentro}}} \text{ ó } \frac{CM_{\text{Filas}}}{CM_{\text{Dentro}}} \tag{13-11}$$

$$F_{\text{Columnas}} = \frac{S^2_{\text{Columnas}}}{S^2_{\text{Dentro}}} \text{ ó } \frac{CM_{\text{Columnas}}}{CM_{\text{Dentro}}} \tag{13-12}$$

$$F_{\text{Interacción}} = \frac{S^2_{\text{Interacción}}}{S^2_{\text{Dentro}}} \text{ ó } \frac{MS_{\text{Interacción}}}{MS_{\text{Dentro}}} \tag{13-13}$$

En Las fórmulas anteriores, F_{Filas} es la razón F del efecto principal de las filas; F_{Columnas} es la razón F del efecto principal de las columna, y $F_{\text{Interacción}}$ es la razón F del efecto interactivo.

Antes de seguir avanzando, es necesario que veamos cómo se calculan los distintos grados de libertad y cómo se diseña la tabla del análisis de varianza.

Grados de libertad en el análisis de varianza de dos criterios

Grados de libertad de las estimaciones intergrupales de varianza a partir de los efectos principales. Los grados de libertad de cada efecto principal (cada estimación intergrupal de varianza) son iguales a la cantidad de niveles de la variable menos 1. Por ejemplo, si existen dos niveles, como sucede en cada efecto principal en el estudio de Lambert et al., hay 1 grado de libertad. En los ejemplos que analizamos anteriormente, relacionados con los niveles de excitación y de dificultad de la tarea, el efecto principal de las columnas (nivel de excitación) tenía tres niveles y, por lo tanto, ese efecto principal tenía 2 grados de libertad.

Lo anterior se expresa bajo las siguientes fórmulas,

$$gl_{\text{Columnas}} = N_{\text{Columans}} - 1 \tag{13-14}$$

y

$$gl_{\text{Filas}} = N_{\text{Filas}} - 1 \tag{13-15}$$

En las fórmulas anteriores, N_{Columnas} = cantidad de columnas, y N_{Filas} = cantidad de hileras.

Grados de libertad de la estimación de la varianza a partir del efecto interactivo. Los grados de libertad de la estimación intergrupal de varianza, calculados a partir del efecto interactivo, son iguales a la cantidad total de casillas menos la cantidad de grados de libertad de los dos efectos principales menos 1. En el estudio de Lambert et al., hay cuatro casillas, y cada efecto principal tiene un grado de libertad. Es decir, quedan 2 grados de libertad a los que se les resta 1 más, y queda 1 para la interacción. En los ejemplos relacionados con el nivel de excitación y dificultad de la tarea, había 6 casillas. El efecto de las columnas tenía 2 grados de libertad, y el efecto de las filas (tarea fácil versus tarea difícil) tenía 1. Es decir, hay 3 grados de libertad, y al restar 1 más quedan 2 grados de libertad para la interacción.

Lo anterior se expresa bajo la siguiente fórmula,

$$gl_{\text{Interacción}} = N_{\text{Casillas}} - gl_{\text{Columnas}} - gl_{\text{Filas}} - 1 \tag{13-16}$$

En la fórmula anterior, N_{Casillas} = cantidad de casillas.

Si aplicamos la fórmula al estudio de Lambert et al.,

$$gl_{\text{Interacción}} = N_{\text{Casillas}} - gl_{\text{Columnas}} - gl_{\text{Fila}} - 1 = 4 - 1 - 1 - 1 = 1$$

Si aplicamos la fórmula al ejemplo acerca de la excitación y la dificultad de la tarea,

$$gl_{\text{Interacción}} = N_{\text{Casillas}} - gl_{\text{Columnas}} - gl_{\text{Fila}} - 1 = 6 - 2 - 1 - 1 = 2$$

Grados de libertad de la estimación intragrupal de varianza poblacional. Como es habitual, los grados de libertad intragrupales son la suma de los grados de libertad de todos los grupos (en este caso, todas las casillas). Tomamos la cantidad de observaciones de cada casilla, le restamos 1 y después sumamos el resultado de todas las casillas. Lo anterior se expresa bajo la siguiente fórmula,

$$gl_{\text{Dentro}} = gl_1 + gl_2 + \ldots + gl_{\text{Último}} \tag{13-17}$$

En la fórmula anterior, $gl_1 + gl_2 + \ldots + gl_{\text{Último}}$ son los grados de libertad de cada casilla (la cantidad de registros de la casilla menos 1) sucesivamente, desde la primera casilla hasta la última.

Grados totales de libertad. Los grados totales de libertad, generalmente son la cantidad de registros menos 1. Se expresa bajo la fórmula,

$$gl_{\text{Total}} = N - 1 \tag{13-18}$$

También podemos calcular los grados totales de libertad sumando todos los grados individuales de libertad (los de las columnas, de las filas, de la interacción y de las intragrupales). Teniendo en cuenta lo anterior, podemos controlar los cálculos aritméticos realizados al calcular los grados de libertad. La fórmula es la siguiente,

$$gl_{\text{Total}} = gl_{\text{Columnas}} + gl_{\text{Filas}} + gl_{\text{Interacción}} + gl_{\text{Dentro}} \tag{13-19}$$

Tabla para un análisis de varianza de dos criterios

La tabla del análisis de varianza de un análisis de dos criterios es similar a la que hemos visto en el capítulo 12 (donde estábamos realizando análisis de varianza de un criterio). Sin embargo, con un análisis de varianza de dos criterios se incluye una línea para cada efecto intergrupal. La tabla 13-5 muestra el diseño.

Tabla 13-5.
Diseño de una tabla de análisis de varianza para un análisis de varianza de dos criterios.

Fuente	*SC*	*gl*	*CM*	*F*
Intergrupal:				
Columnas	$SC_{Columnas}$	$gl_{Columnas}$	$CM_{Columnas}$	$F_{Columnas}$
Filas	SC_{Filas}	gl_{Filas}	CM_{Filas}	F_{Filas}
Interacción	$SC_{Interacción}$	$gl_{Interacción}$	$CM_{Interacción}$	$F_{Interacción}$
Interacciones	SC_{Dentro}	gl_{Dentro}	CM_{Dentro}	
Total	SC_{Total}	gl_{Total}		

Ejemplo de cálculo de un análisis de varianza de dos criterios

Wong y Csikszentmihalyi (1991) realizaron un estudio en el cual, durante una semana, 170 alumnos de escuela portaron equipos buscapersonas y se los llamaba a intervalos aleatorios (aproximadamente cada 2 horas durante el día). Cada vez que recibían una llamada, los alumnos debían llenar un formulando indicando qué estaban haciendo en ese momento. El estudio era un diseño factorial 2 x 2 que analizaba el efecto del sexo y el nivel de los alumnos obtenido en una prueba acerca del deseo de relacionarse. La variable medida era la cantidad de ocasiones, durante la semana, en las que el alumno estaba realizando actividades sociales cuando se lo llamaba. (También había otras variables, pero nos concentraremos sólo en éstas).

Las medias de casilla y marginales de los resultados aparecen en la tabla 13-6, exactamente como fueron informadas por Wong y Csikszentmihalyi. Sin embargo, para que el ejemplo sea simple a los efectos del aprendizaje, hemos inventado valores que producen las mismas medias de casillas y marginales pero incluyen sólo 10 participantes por casilla. La tabla 13-7 incluye estos valores y los cálculos de todos los desvíos.

Tabla 13-6.
Medias de casillas y marginales de la cantidad de veces que se encontró a los participantes en actividades sociales.

		Deseo de relación		
		Bajo	*Alto*	
Sexo	*Muchachos*	10,30	9,22	9,76
	Muchachas	15,75	18,51	17,13
		13,03	13,87	13,45

Fuente: Wong & Csikszentmihalyi, 1991.

Tabla 13-7.
Observaciones, desvíos cuadráticos y sumas de desvíos cuadráticos de los datos ficticios basados en el estudio de Wong y Csikszentmihalyi (1991).

Bajo nivel de deseo de relacionarse						Alto nivel de deseo de relacionarse					
X	$(X-GM)^2$	$(X-M)^2$	$(M_{Fila}-GM)^2$	$(M_{Columna}-GM)^2$	INT^2	X	$(X-GM)^2$	$(X-M)^2$	$(M_{Fila}-GM)^2$	$(M_{Columna}-GM)^2$	INT^2
Muchachos											
12,1	1,82	3,24	13,62	0,18	0,92	11,1	5,52	3,53	13,62	0,18	0,92
11,4	4,20	1,21	13,62	0,18	0,92	10,4	9,30	1,39	13,62	0,18	0,92
11,2	5,06	0,81	13,62	0,18	0,92	10,2	10,56	0,96	13,62	0,18	0,92
10,9	6,50	0,36	13,62	0,18	0,92	9,8	13,32	0,34	13,62	0,18	0,92
10,3	9,92	0,00	13,62	0,18	0,92	9,2	18,06	0,00	13,62	0,18	0,92
9,8	13,32	0,25	13,62	0,18	0,92	9,1	18,92	0,01	13,62	0,18	0,92
9,7	14,06	0,36	13,62	0,18	0,92	8,9	20,70	0,10	13,62	0,18	0,92
9,5	15,60	0,64	13,62	0,18	0,92	8,7	22,56	0,27	13,62	0,18	0,92
9,3	17,22	1,00	13,62	0,18	0,92	8,2	27,56	1,04	13,62	0,18	0,92
8,8	21,62	2,25	13,62	0,18	0,92	6,6	46,92	6,86	13,62	0,18	0,92
103,0	109,32	10,12	136,20	1,80	9,20	92,2	193,42	14,50	136,20	1,80	9,20
Muchachas											
17,4	15,60	2,74	13.54	0,18	0,92	22,0	73,10	2,72	13,54	0,18	0,92
17,1	13,32	1,82	13,54	0,18	0,92	20,5	49,70	3,96	13,54	0,18	0,92
16,8	11,22	1,10	13,54	0,18	0,92	19,9	41,60	1,93	13,54	0,18	0,92
16,7	10,56	0,90	13,54	0,18	0,92	19,1	31,92	0,35	13,54	0,18	0,92
15,5	4,20	0,06	13,54	0,18	0,92	18,5	25,50	0,00	13,54	0,18	0,92
15,3	3,42	0,20	13,54	0,18	0,92	17,4	15,60	1,23	13,54	0,18	0,92
15,0	2,40	0,56	13,54	0,18	0,92	17,0	12,60	2,28	13,54	0,18	0,92
15,4	3,80	0,12	13,54	0,18	0,92	17,1	13,32	1,99	13.54	0,18	0,92
14,3	0,72	2,10	13,54	0,18	0,92	17,1	13,32	1,99	13,54	0,18	0,92
14,0	0,30	3,06	13,54	0,18	0,92	16,5	9,30	4,04	13,54	0,18	0,92
157,5	65,54	12,64	135,40	1,80	9,20	185,1	285,96	29,95	135,40	1,80	9,20

M = media de la casilla de la observación.
M_{fila} = media de la fila de la observación.
$M_{Columna}$ = media de la columna de la observación.
INT = desvío residual de la observación para la interacción

Ejemplo de los cálculos de desvíos utilizando la primera observación en la casilla correspondiente a los muchachos con bajo nivel:

$$(X-GM)^2 = (12{,}1-13{,}45)^2 = -1{,}35^2 = 1{,}82$$
$$(X-M)^2 = (12{,}1-10{,}30)^2 = 1{,}80^2 = 3{,}24$$
$$(M_{Fila}-GM)^2 = (9{,}76-13{,}45)^2 = -3{,}69^2 = 13{,}62$$
$$(M_{Columna}-GM)^2 = (13{,}03-13{,}45)^2 = -0{,}42^2 = 0{,}18$$
$$\text{INT}^2 = [(X-GM)-(X-M)-(M_{Fila}-GM)-(M_{Columna}-GM)]^2 = [(-1{,}35)-(1{,}80)-(-3{,}69)-(-0{,}42)]^2$$
$$= (-1{,}35-1{,}80+3{,}69+0{,}42)^2 = 0{,}96^2 = 0{,}92$$

$$SC_{Total} = 109{,}32 + 193{,}42 + 65{,}54 + 285{,}96 = 654{,}24$$
$$SC_{Dentro} = 10{,}12 + 14{,}50 + 12{,}64 + 29{,}95 = 67{,}21$$
$$SC_{Fila} = 136{,}20 + 136{,}20 + 135{,}40 + 135{,}40 = 543{,}20$$
$$SC_{Columna} = 1{,}80 + 1{,}80 + 1{,}80 + 1{,}80 = 7{,}20$$
$$SC_{Interacción} = 9{,}20 + 9{,}20 + 9{,}20 + 9{,}20 = 36{,}80$$

Control de exactitud: $SC_{Total} = 654{,}24$; $SC_{Dentro} + SC_{Fila} + SC_{Columnas} + SC_{Interacción} = 67{,}21 + 543{,}20 + 7{,}20 + 36{,}80 = 654{,}41$
(Los resultados contemplan diferencias de redondo).

Tabla 13-8.
Cálculo de un análisis de varianza utilizando sumas de cuadrados, sobre la base del estudio de Wong y Csikzentmihalyi (1991). (Datos Ficticios).

F punto de corte para el efecto principal del sexo ($gl = 1, 36$; $p < 0{,}05$) = 4,12 ($gl = 1, 35$ de la tabla)
F punto de corte para el efecto principal de deseo de relacionarse para ($gl = 1, 36$; $p < 0{,}05$) = 4,12
F punto de corte para el efecto interactivo ($gl = 1, 36$; $p < 0{,}05$) = 4,12

Fuente	*SC*	*gl*	*CM*	*F*	
Sexo	543,20	1	543,20	290,48	Se rechaza la hipótesis nula.
Deseo de relacionarse	7,20	1	7,20	3,85	No se rechaza la hipótesis nula.
Sexo por deseo de relacionarse	36,80	1	36,80	19,68	Se rechaza la hipótesis nula.
Dentro de las casillas	67,21	36	1,87		

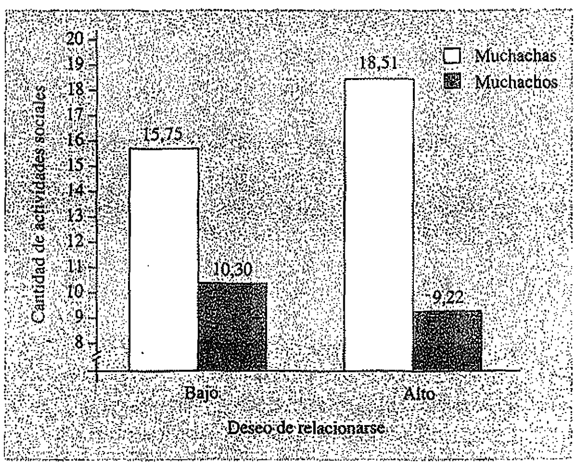
La tabla 13-8 muestra los valores correspondientes al punto de corte F y a la tabla del análisis de varianza. La *figura 13-9 representa los resultados gráficamente.* Analizaremos el ejemplo siguiendo el procedimiento habitual de prueba de hipótesis paso a paso.

1. **Replantear el problema en función de hipótesis de investigación e hipótesis nula de las poblaciones para cada efecto principal y efecto interactivo.** Existen cuatro poblaciones:

Población 1,1: muchachas que tienen un nivel bajo de deseo de relacionarse.
Población 1,2: muchachas que tienen un nivel alto de deseo de relacionarse.
Población 2,1: muchachos que tienen un nivel bajo de deseo de relacionarse.
Población 2,2: muchachos que tienen un nivel alto de deseo de relacionarse.

La primer hipótesis nula establece que las poblaciones combinadas de muchachas (poblaciones 1,1 y 1,2) tienen la misma media que las poblaciones combinadas de muchachos (2,1 y 2,2), en cuanto a la cantidad de veces que se involucraron en actividades sociales. Esta es la hipótesis nula que prueba el efecto principal del sexo (muchachas en contraposición con muchachos). La hipótesis de investigación establece que las poblaciones de muchachas y de muchachos tienen diferentes medias.

Figura 13-9.
Gráfico de los datos ficticios (simplificados) basados en los resultados del estudio de Wong y Csikszentmihalyi (1991).



La segunda hipótesis nula establece que las poblaciones combinadas de aquellos con bajo deseo de relacionarse (poblaciones 1,1 y 2,1) tienen la misma media que las poblaciones combinadas de aquellos con alto deseo de asociación (poblaciones 1,2 y 2,2), con respecto a la cantidad de veces que se involucraron en actividades sociales. Esta es la hipótesis nula que prueba el efecto principal del deseo de relacionarse (bajo en contraposición con alto). La hipótesis de investigación establece que las poblaciones con alto y bajo nivel de deseo de relacionarse tienen diferentes medias.

La tercer hipótesis nula establece que la diferencia entre la cantidad media de actividades sociales de las dos poblaciones de muchachas (población 1,1 menos población 1,2) será la misma que la diferencia entre las medias de las dos poblaciones de muchachos (población 2,1 menos población 2,2). Esta es la hipótesis que prueba el efecto interactivo. (También podría haberse planteado, sin cambiar el significado, como la diferencia entre las dos poblaciones con bajo nivel de deseo, igualando la diferencia entre las dos poblaciones con alto nivel de deseo). La hipótesis de investigación establece que estas diferencias no serán iguales.

2. **Determinar las características de las distribuciones comparativas.** Las tres distribuciones comparativas serán distribuciones *F*. Los grados de libertad de los denominadores serán la suma de los grados de libertad de cada una de las casillas (la cantidad de observaciones de la casilla menos 1). En este caso, hay 10 participantes en cada una de las cuatro casillas, es decir, 9 grados de libertad por casilla; queda un total de 36. El numerador de la distribución comparativa del efecto principal del sexo tiene 1 grado de libertad (2 filas menos 1); el numerador del efecto principal del deseo de relacionarse también tiene 1 grado de libertad, y el grado de libertad del numerador del efecto interactivo es, nuevamente, 1 (es la cantidad de casillas, 4, menos los grados de libertad de las columnas, menos los grados de libertad de las filas, menos 1). Como control de la precisión del cálculo de los grados de libertad, los tres numeradores más los grados de libertad del denominador son igual a $1 + 1 + 1 + 36 = 39$; lo que es igual al total de grados de libertad calculados como la cantidad de participantes menos 1 (es decir, $40 - 1 = 39$).

3. **Determinar los puntos de corte en las distribuciones comparativas, a partir de los cuales debería rechazarse cada hipótesis nula.** Utilizando el nivel 0,05 de significación, la tabla B-3 indica un punto de corte para 1 y 35 grados de libertad de 4,12 (el más cercano disponible en la tabla debajo de 1 y 36). Los grados de libertad y el nivel de significación son los mismos, en este caso, para ambos efectos principales y para el efecto de interacción; por lo tanto, el punto de corte también es el mismo para los tres efectos.

4. **Determinar los valores muestrales en cada distribución comparativa.** Este paso requiere el cálculo de tres razones *F*, que, como hemos visto, requiere calcular primero varios desvíos, elevarlos al cuadrado y sumarlos. La tabla 13-7 indica los desvíos cuadráticos de cada participante.

Para ahorrar espacio, la tabla indica sólo los desvíos cuadráticos. Sin embargo, debajo de la tabla de desvíos cuadráticos mostramos un ejemplo de la forma de cálculo de los desvíos cuadráticos de la primera observación. Dos consejos son especialmente útiles para el cálculo del desvío en el efecto interactivo: a) se debe prestar mucha atención a los signos de los cuadrados que se están restando y b) no se debe olvidar que el desvío interactivo, antes de elevarse al cuadrado, se calcula a partir de los desvíos originales no elevados al cuadrado, y no de los desvíos cuadráticos.

Después los desvíos cuadráticos individuales se suman para obtener SC_{Total}, y así sucesivamente, como lo indica la siguiente parte de la tabla 13-7. Es importante recordar que las **sumas** de los distintos desvíos cuadráticos (SC_{Dentro} ó SC_{Filas}, SC_{Columnas}, $SC_{\text{Interacción}}$) conforman el desvío cuadrático total. Sin embargo, si tomamos a un sólo participante, los distintos desvíos cuadráticos no dan el desvío cuadrático general a la observación con respecto a la gran media. La tabla 13-7 también indica el control de la exactitud de los cálculos: la suma de los desvíos cuadráticos con respecto a la gran media es igual al total de las sumas de las otras cuatro clases de desvíos cuadráticos (teniendo en cuenta las diferencias de redondeo).

Existe otro detalle importante con respecto a los cálculos, que se indican en la tabla 13-7. Comúnmente, en un análisis 2 x 2 todos los desvíos cuadráticos de las filas son iguales (como lo son todos los desvíos cuadráticos de las columnas y todos los desvíos cuadráticos de la interacción). La pequeña diferencia (136,20 contra 135,40) entre los desvíos cuadráticos de las filas de la parte inferior y las de la parte superior se debe simplemente a diferencias de redondeo al calcular las medias de las filas.

Los siguientes pasos se indican en la tabla 13-8 del análisis de varianza. Primero, ingresamos la suma de los desvíos cuadráticos de la tabla anterior para cada estimación de la varianza, y además los grados de libertad del paso 2. Después utilizamos esos datos para calcular el resto de la tabla (los cuadrados medios y los valores *F*). Las conclusiones se indican en el extremo derecho de la tabla y se detallan en el punto 5.

5. **Comparar los valores de los pasos 3 y 4 para decidir si se rechazan o no las hipótesis nulas.** La razón *F* calculada para el efecto principal del sexo, de 290,48, es mucho mayor que el punto de corte de 4,12. Por lo tanto, podemos rechazar la hipótesis nula que establece que las poblaciones de muchachas y muchachos tienen la misma cantidad media de actividad social. Es decir, el efecto principal del sexo es significativo. El *F* de 3,85, correspondiente al efecto principal de la necesidad de relacionarse, no llegó al 4,12, punto de corte necesario. Se puede decir que este efecto se ha acercado a la significación pero no la ha alcanzado. Finalmente, el efecto interactivo *F* de 19,68 excede el punto de corte de 4,12; por lo tanto, el efecto interactivo también es significativo. (En el estudio real se encontró el mismo patrón básico, el efecto principal del sexo y el efecto interactivo eran significativos, mientras que el efecto principal del deseo de relacionarse se acercó pero no llegó a la significación). Antes de continuar, sería una buena idea que el alumno intentara explicar con palabras el significado de esta interacción.

La figura 13-9 representa gráficamente el patrón de medias. Como podemos observar en el gráfico (y por las medias de casillas de la tabla 13-6), el efecto principal del sexo se debe a que las muchachas participan en más actividades sociales que los muchachos. El efecto interactivo se debe a que el deseo de relacionarse está asociado con la mayor cantidad de actividad social de las muchachas, pero básicamente no relacionado con la cantidad de actividades sociales de los muchachos. Es decir, existía una diferencia entre la cantidad de actividades de muchachas con alto nivel y aquellas con bajo nivel de deseo de relacionarse. Pero entre muchachos, la diferencia casi no existía (incluso se daba levemente en la dirección contraria). Esta es la razón por la cual, en general, combinando muchachos y muchachas, el deseo de relacionarse parecía tener poca o ninguna influencia en las actividades. Una vez más podemos observar cómo se descubre una relación interesante entre las variables a través de la aplicación del análisis de varianza para analizar efectos interactivos.

Un segundo ejemplo de cálculo de un análisis de varianza de dos criterios.

Blanchard, Lilly y Vaughn (1991) realizaron un estudio referido a las influencias sobre la expresión de la reacción al racismo. En este estudio se contactaron 72 mujeres blancas no graduadas mientras se dirigían de una clase a otra, y se les pidió que participaran en una encuesta sobre qué tendría que hacer la facultad en respuesta a los anónimos raciales. Se trataba de un diseño 2 x 3.

El factor de dos niveles se dividía en aquellas alumnas a las que se les pedía que respondieran en forma privada (por escrito) y a las que se les pedía que respondieran en forma pública (oralmente al investigador sin que nadie más escuchara). El factor de tres niveles era la dirección de la influencia. Un tercio de las alumnas estaba en condición de "ausencia de influencia", en la que simplemente completaban una encuesta sin nadie presente, excepto el experimentador. Este era un nivel del factor. En el caso de los otros dos tercios de las alumnas, antes de que pudieran comenzar a contestar se invitaba a otra alumna a participar. La otra alumna en realidad era una aliada del experimentador, y se arregló que ella siempre contestara primero. Sus opiniones o eran muy antirracistas o no eran antirracistas en absoluto, creando así los otros dos niveles del factor, "influencia antirracista" e "influencia no antirracista".

Para favorecer la claridad de la explicación, una vez más hemos construido datos que concuerdan con el patrón básico de los descubrimientos reales del estudio, pero que incluye muchos menos participantes. También hemos utilizado valores con números enteros. Los resultados, utilizando estos valores, están representados gráficamente en la figura 13-10. La tabla 13-9 indica las observaciones, los desvíos cuadráticos, los cálculos intermedios y la tabla del análisis de varianza. Seguiremos este ejemplo paso a paso.

1. **Replantear el problema en función de hipótesis de investigación e hipótesis nula de las poblaciones para cada efecto principal e interactivo.** Existen seis poblaciones:

Población 1,1: alumnas que responden públicamente con influencia antirracista.
Población 1,2: alumnas que responden en forma privada con influencia antirracista.
Población 2,1: alumnas que responden públicamente sin influencia.
Población 2,2: alumnas que responden en forma privada sin influencia.
Población 3,1: alumnas que responden públicamente con influencia no antirracista.
Población 3,2: alumnas que responden en forma privada con influencia no antirracista.

La primera hipótesis nula establece que las poblaciones combinadas de alumnas que responden públicamente (poblaciones 1,1, 2,1 y 3,1) tienen la misma media que las poblaciones combinadas de alumnas que responden en forma privada (poblaciones 1,2, 2,2 y 3,2), en cuanto a la expresión de actitudes antirracistas. Esta es la hipótesis nula para probar el efecto principal de la modalidad de la respuesta (pública contra privada). La hipótesis de investigación establece que

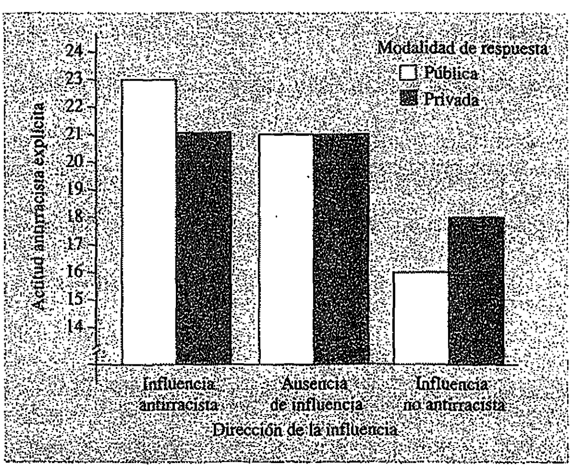


Figura 13-10.
Gráfico de los datos ficticios (simplificados) basados en los resultados del estudio de Blanchard, Lilly y Vaughn.

las poblaciones de alumnas que responden pública y privadamente tienen diferentes medias en cuanto a la expresión de actitudes antirracistas.

La segunda hipótesis nula establece que no existe diferencia entre las medias de las poblaciones combinadas expuestas a influencia antirracista (poblaciones 1,1 y 1,2), las poblaciones combinadas no expuestas a ninguna influencia (poblaciones 2,1 y 2,2) y las poblaciones combinadas expuestas a influencia no antirracista (poblaciones 3,1 y 3,2). Esta es la hipótesis nula que prueba el efecto principal de la dirección de la influencia. La hipótesis de investigación establece que esas tres poblaciones combinadas tienen diferentes medias.

La tercera hipótesis nula establece que el patrón de las medias de las tres poblaciones que responden en público (poblaciones 1,1, 2,1 y 3,1) será igual al patrón de las medias de las tres poblaciones que responden en privado (poblaciones 1,2, 2,2 y 3,2). Esta es la hipótesis nula que prueba el efecto interactivo. (También podría plantearse del siguiente modo sin cambiar el significado: la diferencia entre las poblaciones que responden en público y en privado será la misma al comparar las dos poblaciones con influencia antirracista, las dos poblaciones que no reciben influencia, y las dos poblaciones con influencia no antirracista). La hipótesis de investigación establece que el patrón de las medias de las tres poblaciones que responden públicamente difiere del patrón de las medias de las tres poblaciones que responden en forma privada.

2. **Determinar las características de las distribuciones comparativas.** Las tres distribuciones comparativas serán distribuciones F con grados de libertad del denominador, iguales a la suma de los grados de libertad de cada una de las casillas (la cantidad de observaciones de la casilla menos 1). En este caso, hay 4 observaciones en cada una de las seis casillas, lo que da 3 grados de libertad en cada uno, y un total de 18. El numerador de la distribución comparativa del efecto principal de la modalidad de respuesta tendrá 1 grado de libertad (2 columnas menos 1); el numerador del efecto principal de la dirección de la influencia tendrá 2 grados de libertad (3 filas menos 1), y los grados de libertad del numerador correspondiente al efecto interactivo también será 2 (la cantidad de casillas, 6, menos los grados de libertad de las columnas, 1, menos los grados de libertad de las filas, 2, menos 1). Como control de la exactitud, los grados de libertad de los tres numeradores más los grados de libertad del denominador son igual a 1+2+2+18=23, cantidad que coincide con los grados de libertad totales calculados como la cantidad de participantes menos 1 (24–1=23).

3. **Determinar los puntos de corte en las distribuciones comparativas, a partir de los cuales se debería rechazar cada hipótesis nula.** Utilizando el nivel 0,05 de significación, la tabla B-3 indica los puntos de corte que aparecen justo debajo de las casillas y las medias marginales en la parte superior de la tabla 13-9.

4. **Determinar los valores muestrales en cada distribución comparativa.** Este paso requiere de tres razones F: se calculan todos los desvíos, se los eleva al cuadrado, se los suma y se los divide por los grados de libertad para obtener los cuadrados medios. Finalmente, se calculan las razones de los distintos cuadrados medios intergrupales y cuadrados medios intragrupales. La tabla 13-9 indica todos los cálculos anteriores.

5. **Comparar los valores de los pasos 3 y 4 para decidir si se rechazan o no las hipótesis nulas.** La razón F, que resultó ser 0, para el efecto principal de la modalidad de respuesta, ciertamente no es significativa. (Si bien se basan en resultados reales, los números específicos son inventados. Utilizando información real, sería muy improbable obtener un F exactamente igual a 0). El F calculado en 11,7, para el efecto principal de la dirección de la influencia, es claramente significativo. Excede en mucho el punto de corte de 3,56. Finalmente, la razón F del efecto interactivo igual a 1,7 no alcanzó el punto de corte 3,56 necesario. Por lo tanto, los resultados no son concluyentes para esta hipótesis. La figura 13-10 representa gráficamente el patrón de las medias. Para observar en el gráfico el efecto principal de la dirección de la influencia, debemos comparar

Tabla13-9.
Cálculos del análisis de varianza de la información ficticia basada en los resultados del estudio de Blanchard, Lilly y Vaughn (1991).

Dirección de la influencia	Modalidad de respuesta: *Pública*	*Privada*	
Influencia antirracista	23	21	22
Ausencia de influencia	21	21	21
Influencia no antirracista	16	18	17
	20	20	20

Punto de corte F necesario para el efecto principal de la modalidad de respuesta ($gl = 1, 18$; $p < 0,05$) = 4,41.
Punto de corte F necesario para el efecto principal de la dirección de la influencia ($gl = 2, 18$; $p < 0,05$) = 3,56.
Punto de corte F necesario para el efecto interactivo ($gl = 2, 18$; $p < 0,05$) = 3,56.

Modalidad pública de respuesta

X	$(X-GM)^2$	$(X-M)^2$	$(M_{Fila} - GM)^2$	$(M_{Columna} - GM)^2$	INT
Influencia antirracista					
25	25	4	4	0	1
20	0	9	4	0	1
23	9	0	4	0	1
24	16	1	4	0	1
	50	14	16	0	4
Ausencia de influencia					
22	4	1	1	0	0
19	1	4	1	0	0
22	4	1	1	0	0
21	1	0	1	0	0
	10	6	4	0	0
Influencia no antirracista					
16	16	0	9	0	1
19	1	9	9	0	1
13	49	9	9	0	1
16	16	0	9	0	1
	82	18	36	0	4

Modalidad privada de respuesta

X	$(X-GM)^2$	$(X-M)^2$	$(M_{Fila} - GM)^2$	$(M_{Columna} - GM)^2$	INT^2
Influencia antirracista					
19	1	4	4	0	1
24	16	9	4	0	1
21	1	0	4	0	1
20	0	1	4	0	1
	18	14	16	0	4
Ausencia de influencia					
24	16	9	1	0	0
18	4	9	1	0	0
22	4	1	1	0	0
20	0	1	1	0	0
	24	20	4	0	0
Influencia no antirracista					
18	4	0	9	0	1
21	1	9	9	0	1
16	16	4	9	0	1
17	9	1	9	0	1
	30	14	36	0	4

M = media de la casilla de la observación
M_{Fila} = media de la fila de la observación
$M_{Columna}$ = media de la columna de la observación
INT = desvío residual de la observación para la interacción

Ejemplo de cálculo de los desvíos, utilizando la primera observación en la casilla pública antirracista:

$$(X-GM)^2 = (25-20)^2 = 5^2 = 25.$$
$$(X-M)^2 = (25-23)^2 = 2^2 = 4.$$
$$(M_{Fila} - GM)^2 = (22-20)^2 = 2^2 = 4.$$
$$(M_{Columna} - GM)^2 = (20-20)^2 = 0^2 = 0.$$
$$INT^2 = [(X-GM) - (X-M) - (M_{Fila} - GM) - (M_{Columna} - GM)]^2 = (5-2-2-0)^2 = 1^2 = 1$$

$$SC_{Total} = 50 + 18 + 10 + 24 + 82 + 30 = 214$$
$$SC_{Dentro} = 14 + 14 + 6 + 20 + 18 + 14 = 86$$
$$SC_{Columna} = 0 + 0 + 0 + 0 + 0 + 0 = 0$$
$$SC_{Fila} = 16 + 16 + 4 + 4 + 36 + 36 = 112$$
$$SC_{Interacción} = 4 + 4 + 0 + 0 + 4 + 4 = 16$$

Control de exactitud: $SC_{Total} = 214$; $SC_{Dentro} + SC_{Fila} + SC_{Columna} + SC_{Interacción} = 86 + 0 + 112 + 16 = 214$

Fuente	SC	gl	CM	F	
Modalidad de respuesta (columnas)	0	1	0	0	No se rechaza la hipótesis nula.
Dirección de la influencia (filas)	112	2	56	11,7	Se rechaza la hipótesis nula.
Interacción (columnas × filas)	16	2	8	1,7	No se rechaza la hipótesis nula.
Dentro de las casillas	86	18	4,8		

las alturas promedio de los pares de barras. La comparación sugiere que este efecto principal se debe a que el grupo que recibió influencia antirracista expresó las actitudes más antirracistas; el grupo que recibió influencia sin dirección expresó un grado intermedio de actitud antirracista, y el grupo que recibió influencia no antirracista expresó actitudes antirracistas mucho menores. También podemos observar que el patrón de las barras es diferente dentro de cada par, sugiriendo un posible efecto interactivo. Sin embargo, la interacción no fue lo suficientemente fuerte en este estudio como para ser considerada significativa. (Es raro encontrar patrones de barras completamente idénticos en estos gráficos, del mismo modo que es raro obtener medias exactamente iguales). Si analizamos el patrón de medias en la parte superior de la tabla 13-9 notaremos el mismo patrón de resultados.

Resumen de los procedimientos para realizar un análisis de varianza de dos criterios

La tabla 13-10 resume los pasos de la prueba de hipótesis, y la tabla 13-11 muestra la tabla del análisis de varianza y las fórmulas para un análisis de varianza de dos criterios. Como ya hemos visto en capítulos anteriores, con respecto a otros procedimientos también existen fórmulas de cálculo de las sumas de los cuadrados que los investigadores utilizaban antes del advenimiento de la computadora. Estas fórmulas facilitan la realización manual de los cálculos (o con una calculadora) cuando se analizan resultados de un estudio real con una gran cantidad de observaciones en cada casilla. En el apéndice I, al final del capítulo, se pueden encontrar las fórmulas de cálculo tradicionales y un ejemplo solucionado con dichas fórmulas. Sin embargo, una vez más recomendamos *insistentemente* que al resolver los ejercicios del libro, se utilicen las fórmulas de definición y los procedimientos que aparecen en las tablas 13-10 y 13-11.

Las fórmulas de definición a las que nos referimos arriba refuerzan los principios implícitos que son, en definitiva, los principales elementos que el alumno está intentando aprender.

Supuestos del análisis de varianza de dos criterios

Los supuestos de un análisis factorial de varianza son, en su mayoría, los mismos que los del análisis de varianza de un criterio que analizamos en el capítulo 11. Del mismo modo, las cuestiones relacionadas con los efectos producidos por el incumplimiento de esos supuestos también son, en gran medida, las mismas que las del análisis de varianza de un criterio. (De hecho, como observamos en el capítulo 11, esos supuestos y cuestiones también son, en gran parte, los mismos que los de la prueba *t* para medias independientes). No obstante, con el análisis factorial de varianza, los supuestos de normalidad de población e igualdad de varianzas se aplican a las poblaciones que corresponden a cada casilla.

POTENCIA Y TAMAÑO DEL EFECTO EN EL ANÁLISIS FACTORIAL DE VARIANZA

En un análisis factorial de varianza calculamos la potencia y el tamaño del efecto casi del mismo modo que en el análisis de varianza de un criterio (*véase* capítulos 11 y 12). La diferencia principal radica en que ambos se calculan separadamente para cada efecto principal e interactivo. Es muy posible, por ejemplo, que la potencia de uno de los efectos principales sea mucho mayor que la del otro efecto principal. Del mismo modo, es posible que el efecto interactivo tenga más o menos potencia que los efectos principales. Es decir, que para planificar un estudio de forma segura

Tabla 13-10.
Pasos de la prueba de hipótesis para el análisis de varianza de dos criterios.

1. Replantear el problema en función de hipótesis de investigación e hipótesis nula de las poblaciones para cada efecto principal y efecto interactivo.
2. Determinar las características de las distribuciones comparativas.
   a) Los grados de libertad del numerador de la distribución F del efecto principal de las columnas son la cantidad de columnas menos 1: $gl_{\text{Columnas}} = N_{\text{Columnas}} - 1$.
   b) Los grados de libertad del numerador de la distribución F del efecto principal de las filas son la cantidad de filas menos 1: $gl_{\text{Filas}} = N_{\text{Filas}} - 1$.
   c) Los grados de libertad del numerador de la distribución F del efecto interactivo son la cantidad de casillas menos los grados de libertad de las columnas, menos los grados de libertad de las filas, menos 1: $gl_{\text{Interacción}} = N_{\text{Casillas}} - gl_{\text{Columnas}} - gl_{\text{Filas}} - 1$.
   d) Las distribuciones comparativas serán distribuciones F con grados de libertad del denominador, igual a la suma de los grados de libertad de cada casilla (la cantidad de casos en la casilla menos 1): $gl_{\text{Dentro}} = gl_1 + gl_2 + \ldots + gl_{\text{Último}}$.
   e) Se controla la exactitud de los cálculos asegurándose de que todos los grados de libertad sumen los grados de libertad totales: $gl_{\text{Total}} = N - 1 = gl_{\text{Dentro}} + gl_{\text{Columnas}} + gl_{\text{Filas}} + gl_{\text{Interacción}}$.
3. Determinar los cortes muestrales en las distribuciones comparativas, a partir de los cuales se debería rechazar cada hipótesis nula.
   a) Determinar los niveles de significación deseados.
   b) Buscar los puntos de corte adecuados en una tabla F (tabla B-3).
4. Determinar los valores muestrales en cada distribución comparativa (serán razones F).
   a) Calcular la media de cada casilla, fila y columna más la gran media de todas las observaciones.
   b) Calcular los siguientes desvíos de cada observación.
      i) El desvío con respecto a la gran media: $X - GM$.
      ii) El desvío con respecto a la media de su casilla $X - M$.
      iii) El desvío de la media de su fila con respecto a la gran media: $M_{\text{Fila}} - GM$.
      iv) El desvío de la media de su columna con respecto a la gran media: $M_{\text{Columna}} - GM$.
      v) El desvío con respecto a la gran media menos todos los otros desvíos: desvío Interactivo = $(X - GM) - (X - M) - (M_{\text{Fila}} - GM)\text{-}(M_{\text{Columna}} - GM)$. (Para calcular este desvío es necesario asegurarse de utilizar los desvíos no cuadráticos y de prestar mucha atención a los signos).
   c) Elevar cada uno de los desvíos al cuadrado.
   d) Calcular las sumas de cada uno de los distintos tipos de desvíos cuadráticos (SC_{Total}, SC_{Dentro}, SC_{Columna}, SC_{Fila} y $SC_{\text{Interacción}}$).
   e) Controlar la exactitud de los cálculos asegurándose de que la suma de los desvíos cuadráticos, basados en el desvío de cada observación con respecto a la gran media, sea igual a la suma de todas las otras sumas de desvíos cuadráticos. $SC_{\text{Total}} = SC_{\text{Dentro}} + SC_{\text{Columna}} + SC_{\text{Fila}} + SC_{\text{Interacción}}$.
   f) Calcular la estimación intergrupal de varianza para cada efecto principal e interactivo (CM_{Columnas} ó $S^2_{\text{Columnas}} = SC_{\text{Columnas}}/gl_{\text{Columna}}$; CM_{Filas} ó $S^2_{\text{Filas}} = SC_{\text{Filas}}/gl_{\text{Filas}}$; $CM_{\text{Interacción}}$ ó $S^2_{\text{Interacción}} = SC_{\text{Interacción}}/gl_{\text{Interacción}}$).
   g) Calcular la estimación intragrupal de varianza (CM_{Dentro} ó $S^2_{\text{Dentro}} = SC_{\text{Dentro}}/gl_{\text{Dentro}}$).
   h) Calcular las razones F para cada efecto principal e interactivo ($F_{\text{Columnas}} = S^2_{\text{Columnas}}/S^2_{\text{Dentro}}$ ó $CM_{\text{Columnas}}/CM_{\text{Dentro}}$; $F_{\text{Fila}} = S^2_{\text{Fila}}/S^2_{\text{Dentro}}$ ó $CM_{\text{Fila}}/CM_{\text{Dentro}}$; $F_{\text{Interacción}} = S^2_{\text{Interacción}}/S^2_{\text{Dentro}}$ ó $CM_{\text{Interacción}}/CM_{\text{Dentro}}$.
5. Comparar los valores obtenidos en los pasos 3 y 4 para decidir si se rechazan o no las hipótesis nulas.

Tabla 13-11.
Tabla del análisis de varianza y fórmulas del análisis de varianza en dos sentidos.

Tabla de análisis de varianza:

Fuente	SC	gl	CM	F
Intergrupal:				
Columnas	SC_{Columnas}	gl_{Columnas}	CM_{Columnas} (ó S^2_{Columnas})	F_{Columnas}
Filas	SC_{Filas}	gl_{Filas}	CM_{Filas} (ó S^2_{Filas})	F_{Filas}
Interacción	$SC_{\text{Interacción}}$	$gl_{\text{Interacción}}$	$CM_{\text{Interacción}}$ (ó $S^2_{\text{Interacción}}$)	$F_{\text{Interacción}}$
Dentro	SC_{Dentro}	gl_{Dentro}	CM_{Dentro} (ó S^2_{Dentro})	
Total	SC_{Total}	gl_{Total}	CM_{Total} (ó S^2_{Total})	

Fórmulas para cada sección de la tabla de análisis de varianza:

Fuente	SC	gl	CM	F
Intergrupal:				
Columnas	$\Sigma(M_{\text{Columna}} - GM)^2$	$N_{\text{Columnas}} - 1$	$SC_{\text{Columnas}}/gl_{\text{Columnas}}$	$CM_{\text{Columnas}}/CM_{\text{Dentro}}$
Filas	$\Sigma(M_{\text{Fila}} - GM)^2$	$N_{\text{Filas}} - 1$	$SC_{\text{Filas}}/gl_{\text{Filas}}$	$CM_{\text{Filas}}/CM_{\text{Dentro}}$
Interacción	$\Sigma[(X - GM) - (X - M) - (M_{\text{Fila}} - GM) - (M_{\text{Columna}} - GM)]^2$	$N_{\text{Casillas}} - N_{\text{Columnas}} - N_{\text{Filas}} - 1$	$SC_{\text{Interacción}}/gl_{\text{Interacción}}$	$CM_{\text{Interacción}}/CM_{\text{Dentro}}$
Dentro	$\Sigma(X - M)^2$	$gl_1 + gl_2 + \ldots + gl_{\text{Último}}$	$SC_{\text{Dentro}}/gl_{\text{Dentro}}$	
Total	$\Sigma(X - GM)^2$	$N - 1$		

Definiciones de los símbolos básicos:

$\Sigma =$	suma de los números correspondientes de todos los casos (no casillas).
$M =$	media de la casilla de la observación.
$M_{\text{Fila}} =$	media de la fila de la observación.
$M_{\text{Columna}} =$	media de la columna de la observación.
$GM =$	gran media de todas las observaciones.
$N_{\text{Casillas}} =$	cantidad de casillas.
$N_{\text{Filas}} =$	cantidad de filas.
$N_{\text{Casillas}} =$	cantidad de columnas.
$X =$	cada observación.
$N =$	cantidad total de casos en el estudio.

necesitamos contar con suficientes participantes como para que el efecto con el menor tamaño del efecto esperado tenga una potencia adecuada. Del mismo modo, al evaluar los resultados de un experimento factorial, debemos tener en cuenta el tamaño del efecto de cada efecto principal e interactivo separadamente.

Tamaño del efecto

El tamaño del efecto de cada efecto principal e interactivo se puede calcular como f (tal como lo hicimos en el capítulo 11) o como R^2, la proporción de varianza explicada (tal como lo hicimos en el capítulo 12). (No debemos olvidar que R^2 también se denomina frecuentemente eta^2). Sin embargo, en el análisis de varianza factorial, y especialmente cuando se utiliza el método del modelo estructural como lo hemos hecho en este capítulo, es más sencillo trabajar con la proporción de varianza explicada adaptando levemente el procedimiento aprendido en el capítulo 12.

En el capítulo 12 describimos la proporción de varianza explicada como la proporción de los desvíos cuadráticos de las observaciones con respecto a la gran media, que estaba explicada por los desvíos de las medias grupales con relación a la gran media. En un análisis de varianza de un criterio, $R^2 = SC_{\text{Entre}}/SC_{\text{Total}}$.

Analicemos ahora el caso del efecto de las columnas en un análisis de varianza de dos criterios. Ciertamente, podemos sustituir SC_{Entre} por SC_{Columnas}. Tiene sentido considerar al numerador de la proporción mencionada como la suma de los desvíos cuadráticos de las medias de las columnas con respecto a la gran media, es decir, la varianza creada por el efecto de la variable relacionada con las columnas, y no que sea explicada de otro modo.

Sin embargo, ¿qué sucede con el denominador, el total de la varianza a ser explicada en parte por la varianza debido a las columnas? En un análisis de dos criterios, los desvíos cuadráticos de cada observación con respecto a la gran media ahora están parcialmente explicados por los efectos de las filas y los aspectos interactivos, al igual que por el efecto de las columnas. Pero cuando evaluamos la proporción de varianza explicada por el efecto de las columnas, no nos interesa lo que realiza el efecto de las filas o el efecto interactivo. No debe responsabilizarse al efecto de las columnas por la varianza que ya es explicada por el efecto de las filas y el efecto interactivo. Por lo tanto, los desvíos cuadráticos a ser explicados por las columnas deberían incluir sólo aquellos desvíos cuadráticos todavía no explicados por las filas y la interacción. Para expresarlo en fórmulas,

$$R^2_{\text{Columnas}} = \frac{SC_{\text{Columnas}}}{SC_{\text{Total}} - SC_{\text{Filas}} - SC_{\text{Interacción}}} \tag{13-20}$$

Simplemente,

$$R^2_{\text{Filas}} = \frac{SC_{\text{Filas}}}{SC_{\text{Total}} - SC_{\text{Columnas}} - SC_{\text{Interacción}}} \tag{13-21}$$

$$R^2_{\text{Interacción}} = \frac{SC_{\text{Interacción}}}{SC_{\text{Total}} - SC_{\text{Columnas}} - SS_{\text{Filas}}} \tag{13-22}$$

Técnicamente, cada uno de los cálculos anteriores es una R^2 "parcial", porque describen la proporción de varianza explicada por un efecto después de "excluir" los otros efectos. (Volveremos a tratar el tema de correlaciones parciales en el capítulo 17).

En el ejemplo basado en el estudio realizado con los equipos de radio llamadas por Wong y Csikszentmihalyi, R^2 se hubiera calculado de la siguiente forma:

$$\begin{aligned} R^2_{\text{Columnas}}(\text{Deseo de realacionarse}) &= \frac{SC_{\text{Columnas}}}{SC_{\text{Total}} - SC_{\text{Filas}} - SC_{\text{Interacción}}} \\ &= \frac{7{,}20}{654{,}24 - 543{,}20 - 36{,}80} = \frac{7{,}20}{74{,}24} = 0{,}10 \end{aligned}$$

$$\begin{aligned} R^2_{\text{Filas}}\ (\text{Sexo}) &= \frac{SC_{\text{Filas}}}{SC_{\text{Total}} - SC_{\text{Columnas}} - SC_{\text{Interacción}}} \\ &= \frac{543{,}20}{654{,}24 - 7{,}20 - 36{,}80} = \frac{543{,}20}{610{,}24} = 0{,}89 \end{aligned}$$

$$R^2_{\text{Interacción}}\ (\text{Interacción}) = \frac{SC_{\text{Interacción}}}{SC_{\text{Total}} - SC_{\text{Columnas}} - SC_{\text{Filas}}}$$
$$= \frac{36,80}{654,24 - 7,20 - 543,20} = \frac{36,80}{103,84} = 0,35$$

Sobre la base de las reglas de Cohen para R^2 en el análisis de varianza, según las presentamos en el capítulo 12, existe un enorme tamaño del efecto, es decir, una R^2 alta para el sexo y además un buen tamaño del efecto para la interacción. El efecto no significativo del deseo de relacionarse tenía un tamaño del efecto entre mediano y grande. (En el estudio real, los tamaños del efecto eran mucho más pequeños. En el ejemplo, los tamaños del efecto son tan grandes porque inventamos datos con mucha menos varianza que en el estudio real, con el fin de que se pudieran observar los patrones con claridad).

Si un estudio sólo proporciona los valores F y los grados de libertad, se aplica la fórmula que vimos en el capítulo 12, $R^2 = (F)(gl_{\text{Entre}})/[(F)(gl_{\text{Entre}})+gl_{\text{Dentro}}]$, sustituyendo los F y los grados de libertad del efecto correspondiente.

$$R^2_{\text{Columnas}} = \frac{(F_{\text{Columnas}})(gl_{\text{Columnas}})}{(F_{\text{Columnas}})(gl_{\text{Columnas}}) + gl_{\text{Dentro}}} \tag{13-23}$$

$$R^2_{\text{Filas}} = \frac{(F_{\text{Filas}})(gl_{\text{Filas}})}{(F_{\text{Filas}})(gl_{\text{Filas}}) + gl_{\text{Dentro}}} \tag{13-24}$$

$$R^2_{\text{Interacción}} = \frac{(F_{\text{Interacción}})(gl_{\text{Interacción}})}{(F_{\text{Interacción}})(gl_{\text{Interacción}}) + gl_{\text{Dentro}}} \tag{13-25}$$

Por ejemplo, en el estudio realizado con los equipos de radio llamada, el deseo de relacionarse era el efecto principal de las columnas. Calculamos F_{Columnas} en 3,85, los grados de libertad para este efecto (gl_{Columnas}) en 1, y los grados de libertad intragrupales de las casillas (gl_{Dentro}) en 36. Por lo tanto,

$$R^2_{\text{Columnas}} = \frac{(F_{\text{Columnas}})(gl_{\text{Columnas}})}{(F_{\text{Columnas}})(gl_{\text{Columnas}}) + gl_{\text{Dentro}}}$$
$$= \frac{(3,85)(1)}{(3,85)(1)+36} = \frac{3,85}{39,85} = 0,10$$

Potencia

En un análisis factorial de varianza, la potencia de cada efecto está influenciada por el diseño general. Por ejemplo, un efecto de columnas de tres niveles, tendrá distinta potencia si se cruza con un efecto de filas de dos niveles o si se cruza con un efecto de filas de tres niveles. Por lo tanto, el análisis de la potencia es diferente según la cantidad de niveles de un efecto y, teniendo en cuenta cada cantidad de niveles, según la cantidad de niveles con que se cruce.

Para simplificar las cosas, veremos las cifras relacionadas con la potencia sólo para las tres situaciones más comunes del análisis de varianza de dos criterios: todos los efectos en un diseño 2 x 2, un efecto principal de dos niveles (dos filas o dos columnas) en un diseño 2 x 3 y un efecto principal de tres niveles (tres filas o tres columnas) en un diseño 2 x 3. (La potencia de la interacción en un diseño 2 x 3 es la misma que la del efecto principal de tres niveles). La tabla 13-12 indica la potencia aproximada al nivel 0,05 de significación para cada una de las situaciones mencionadas, con tamaños del efecto pequeños, medianos y grandes, y con tamaños de casillas de 10, 20, 30, 40, 50 y 100.[3]

Analicemos un estudio planificado 2 x 2 con 30 participantes por casilla y con un tamaño mediano del efecto esperado ($R^2 = 0{,}06$), a realizarse con el nivel 0,05. El estudio que mencionamos arriba tendría una potencia de 0,78, es decir, que si la hipótesis de investigación en efecto es verdadera y tiene un tamaño del efecto mediano, las posibilidades de que el estudio resulte significativo es de aproximadamente el 78%. O veamos un ejemplo tomado de una publicación, en el que se encontró un resultado no significativo para un efecto interactivo en un análisis de varianza 2 x 3 con 20 participantes por casilla. Basándonos en la tabla, para un tamaño del efecto pequeño el poder del estudio es de sólo 0,14. Es decir, que aun si ese efecto pequeño existe en la población, sería muy improbable que el estudio resultara significativo. Por el contrario, la tabla muestra una potencia de 0,98 para un gran tamaño del efecto; por lo tanto, si existiera un gran efecto en la población, casi con seguridad hubiera resultado significativo en el estudio.

Tabla 13-12.
Potencia aproximada de estudios realizados con un análisis de varianza 2 x 2 ó 2 x 3 con respecto a hipótesis probadas a un nivel 0,05 de significación.

N por Casilla	Tamaño del efecto		
	Pequeño ($f = 0{,}10$) ($R = 0{,}10$) ($R^2 = 0{,}01$)	*Mediano* ($f = 0{,}25$) ($R = 0{,}24$) ($R^2 = 0{,}06$)	*Grande* ($f = 0{,}40$) ($R = 0{,}37$) ($R^2 = 0{,}14$)
Todos los efectos en un análisis 2x2:			
10	0,09	0,33	0,68
20	0,13	0,60	0,94
30	0,19	0,78	0,99
40	0,24	0,89	*
50	0,29	0,94	*
100	0,52	*	*
Efecto principal de dos niveles en un análisis 2x3:			
10	0,11	0,46	0,84
20	0,18	0,77	0,99
30	0,26	0,92	*
40	0,34	0,97	*
50	0,41	0,99	*
100	0,70	*	*
Efecto de tres niveles y efecto interactivo en un análisis 2x3:			
10	0,09	0,36	0,76
20	0,14	0,67	0,98
30	0,21	0,86	*
40	0,27	0,94	*
50	0,32	0,98	*
100	0,59	*	*

*Casi 1.

[3] Cohen (1988, p.389-354) proporciona tablas más detalladas. Sin embargo, utilizar esas tablas con el diseño factorial requiere algunos cálculos preliminares, tal como lo explica Cohen en las páginas 364-379.

Planificación del tamaño de la muestra

La tabla 13-13 indica la cantidad aproximada de participantes necesarios por casilla para obtener una potencia del 80% a un nivel 0,05 de significación, con tamaños del efecto estimados pequeños, medianos y grandes y para los mismos casos incluidos en la tabla de potencia.[4]

Supongamos que queremos planificar un análisis de varianza 2 x 3, en el que predecimos un gran tamaño del efecto para el efecto principal en la variable de tres niveles, y un tamaño del efecto mediano para los otros efectos principales y para el efecto interactivo. Para obtener una potencia del 80% (al nivel 0,05 de significación), necesitaríamos 11 participantes por casilla para el efecto principal de tres niveles, 22 por casilla para el efecto principal de dos niveles y 27 por casilla para el efecto interactivo. Dado que el experimento se realiza completo de una sola vez, debemos tener al menos 27 participantes por casilla (a menos que decidamos arriesgarnos a tener menor potencia para el efecto interactivo). En consecuencia, deberíamos reclutar 162 participantes (27 para cada una de las seis casillas del diseño 2 x 3).

EXTENSIONES Y CASOS ESPECIALES DEL ANÁLISIS FACTORIAL DE VARIANZA

El análisis de varianza es una técnica extremadamente versátil. Con el análisis de varianza de un criterio o de dos criterios, tal como los hemos descripto aquí y en los capítulos 11 y 12, se pueden analizar muchas situaciones de investigación. Existen otras técnicas que son más complejas o en las que hay que tener en cuenta aspectos especiales. No podemos, en este libro introductorio, entrar en detalles de los procedimientos estadísticos necesarios para manejar todas las posibilidades. (Esos procedimientos se incluyen en la mayoría de los textos de estadística de nivel intermedio para la psicología, como también en los que se suelen denominar libros de "diseño experimental"). Sin embargo, podemos describir algunas de las variaciones o aspectos a tener en cuenta, de modo de brindar una idea de las modificaciones básicas que deben realizarse a los procedimientos ya aprendidos.

Tabla 13-13.
Cantidad aproximada de participantes necesarios en cada casilla (suponiendo igual tamaño de muestras) para obtener una potencia del 80% en estudios que utilizan el análisis de varianza 2 x 2 ó 2 x 3, probando las hipótesis a un nivel 0,05 de significación.

	Tamaño del efecto		
	Pequeño ($f = 0,10$) ($R = 0,10$) ($R^2 = 0,01$)	*Mediano* ($f = 0,25$) ($R = 0,24$) ($R^2 = 0,06$)	*Grande* ($f = 0,40$) ($R = 0,37$) ($R^2 = 0,14$)
2 x 2: todos los efectos	197	33	14
2 x 3: efecto principal en dos niveles	132	22	9
efecto principal en tres niveles y efecto interactivo	162	27	11

[4] Cohen (1988, pp. 381-389) proporciona tablas más detalladas. Para utilizarlas es indispensable leer primero las páginas 396-403 de Cohen.

Diseños de análisis de varianza de tres criterios o más

La extensión más directa del análisis de varianza de dos criterios es la de los experimentos que incluyen diseños de tres criterios de clasificación o más. En esos casos, el análisis se realiza exactamente como lo hemos descripto en este capítulo, excepto por la existencia de efectos principales e interactivos adicionales.

A veces un experimento incluye variables que sólo son de interés si interactúan con las variables más importantes. Ejemplos de tales variables son el orden de presentación o cuál de dos experimentadores realizaron el estudio con cada participante. En los casos que acabamos de mencionar, el investigador puede comenzar con un análisis de varianza factorial de varios criterios. Si las variables de interés secundario no tienen efectos interactivos significativos con las variables de interés primario, se vuelve a realizar el análisis ignorando las variables secundarias. El diseño se convierte entonces en un análisis de varianza más manejable, de dos o tres criterios; se dice que el análisis resultante se ha **plegado** sobre las variables que se están ignorando. Por ejemplo, en el estudio de Lambert et al. acerca del estado de ánimo y de la calidad del estereotipo, los investigadores primero realizaron un análisis de tres criterios que incluía el sexo. (Anteriormente en este capítulo, figura 13-1, incluimos un diagrama del caso anterior). Cuando descubrieron que no existían efectos principales ni interactivos en relación con el sexo procedieron a realizar un análisis de dos criterios, plegado sobre el sexo (es decir, el sexo no se tuvo en cuenta en el análisis en dos sentidos).

Análisis de varianza de medidas repetidas

En todas las situaciones que hemos analizado en este capítulo y en los capítulos 11 y 12, las diferentes casillas o agrupaciones se basan en valores de diferentes individuos. A veces, sin embargo, un investigador mide al mismo individuo en varias situaciones diferentes. (Si existen sólo dos situaciones diferentes, como por ejemplo antes y después del tratamiento, podemos utilizar una prueba *t* para medias dependientes, como la descripta en el capítulo 9). Analicemos un estudio en el que el investigador mide la velocidad de reconocimiento de una sílaba cuando está inserta en tres tipos de palabras: palabras familiares, palabras no familiares, y sonidos que no forman palabras. En un estudio de este tipo, es común que se exponga a cada participante a una gran cantidad de palabras de cada tipo ordenadas al azar. Como resultado de ese procedimiento se obtiene una cantidad promedio de errores por cada participante, para cada clase de palabra. Otro tipo de ejemplo es el de un estudio acerca de los efectos de la psicoterapia, en el que se mide la depresión de los pacientes antes, inmediatamente después, y nuevamente a los 3 meses después de haber asistido a terapia. En los dos ejemplos anteriores, tenemos tres grupos de valores observados, sin embargo, cada serie de tres observaciones pertenece a la misma persona. En el ejemplo de las distintas clases de palabras, cada participante tiene una observación (tiempo promedio de respuesta) hecha por cada una de las tres clases de palabras; en el ejemplo de la psicoterapia, cada individuo tiene una observación hecha (nivel de depresión) por cada una de las tres ocasiones (antes, inmediatamente después, 3 meses después).

Los estudios mencionados en los párrafos anteriores son ejemplos de diseños de medidas repetidas. Los diseños de medidas repetidas son analizados con un **análisis de varianza de medidas repetidas**. El nombre se debe a que los mismos participantes son medidos repetidas veces. Este tipo de diseño y análisis también se denomina diseño intra-sujeto y análisis de varianza intra-sujeto, porque la comparación se realiza con respecto al interior de los diferentes participantes o sujetos, y no entre ellos. (La Asociación Americana de Psicología recomienda la utilización del término "participante" más que el de "sujeto").

A veces, una variable de medidas repetidas se cruza en el mismo estudio con una variable común entre participantes. Por ejemplo, en el estudio acerca de la terapia podría haber un grupo de control que no hiciera terapia pero que fuera probado en las mismas tres ocasiones que los demás. En ese caso se estaría realizando un diseño 2 (grupo de terapia contra grupo de control) x 3 (antes, después, 3 meses después), en el que la primera variable es del tipo usual entre participantes, como las que hemos estado utilizando hasta ahora, y la segunda es una variable del tipo de medidas repetidas. Incluso es posible tener dos factores de medidas repetidas o combinaciones aún más complejas.

Los diseños experimentales que incluyen una o más variables de medidas repetidas son bastante comunes, pero son controvertidos en cuanto a la forma en que deberían analizarse. Una de esas formas involucra una extensión bastante directa de los procedimientos que hemos aprendido. El apéndice II de este capítulo describe esos procedimientos para un análisis de varianza de medidas repetidas de un criterio. Sin embargo, los supuestos que deben cumplirse para que este método dé resultados precisos son bastante exigentes y, a menudo, no se cumplen en la práctica. Por eso, algunos investigadores han sostenido enérgicamente que los diseños de medidas repetidas usualmente deberían analizarse utilizando un procedimiento mucho más complejo denominado "*análisis de varianza multivariable*". Otros investigadores defienden el procedimiento normal modificado, pero realizan una adaptación a los grados de libertad utilizados para determinar el punto F de corte.

El mejor método, en términos tanto del error Tipo I como del error Tipo II, sigue siendo controvertido y complejo (el tema ha sido tratado recientemente por Keselman, Lix & Keselman, 1996; Algina & Keselman, 1997). Al igual que con otros temas controvertidos que hemos tratado, en la mayoría de las situaciones reales de investigación los resultados no difieren demasiado con los distintos método utilizados; sin embargo, si los resultados son muy ajustados, es aconsejable no considerarlos tan concluyentes sin realizar una mayor investigación.

CONTROVERSIAS, LIMITACIONES Y DESARROLLOS RECIENTES

En esta sección veremos dos antiguas controversias con respecto al análisis factorial de varianza. Una cuestión trata sobre lo que se debe hacer cuando hay distinta cantidad de participantes en las diversas casillas. La otra cuestión trata sobre el modo de manejar una situación en la que una de las variables no es categórica sino numérica cuantitativa, situación generalmente denominada "dicotomización" de una variable.

Cantidad desigual de participantes en las casillas

¿Qué sucede cuando realizamos un análisis factorial de varianza del modo que hemos descripto y en el cual las casillas no tienen la misma cantidad de participantes? En general, aun utilizando el método del modelo estructural, se obtienen resultados distorsionados. (El análisis de varianza de **un criterio**, con el método de modelo estructural que describimos en el capítulo 12, sí funciona correctamente tanto con tamaños de casillas iguales o desiguales. La situación es más complicada, sin embargo, con un análisis de varianza de dos criterios o más).

Ha habido mucha controversia con respecto al tema mencionado en el párrafo anterior. Un método al que a veces recurren investigadores desalentados consiste en eliminar registros al azar de aquellas casillas que tienen demasiados, pero este método desperdicia potencia. Actualmente se considera en forma generalizada que la solución denominada **análisis de varianza de cuadra-**

dos mínimos es el método óptimo, y la mayoría de los programas para computadora que calculan el análisis de varianza disponen de esta opción; algunos incluso la utilizan automáticamente, a menos que se les indique lo contrario. (El método mencionado se basa en el análisis de regresiones múltiples, del cual hemos aprendido algo en el capítulo 4). Cuando los tamaños de las casillas son iguales, este método da el mismo resultado que el método ordinario.

El resultado de utilizar el método de cuadrados mínimos es que se empareja la influencia de cada casilla sobre los efectos principales e interactivos, que es en realidad lo que se busca. Sin embargo, un documento influyente (Milligan, Wong & Thompson, 1987) sugirió que este método es especialmente susceptible al incumplimiento de los supuestos de normalidad poblacional o de igualdad de las varianzas poblacionales. (Lamentablemente, otros métodos tradicionales del análisis factorial de varianza con tamaños desiguales de casillas son exactamente tan susceptibles como el anterior). Por lo tanto, la mejor recomendación para los investigadores es diseñar estudios que utilicen casillas del mismo tamaño. Además, al igual que sucedía con la prueba *t*, para una determinada cantidad de participantes la potencia es mayor cuando se conforman con ellos grupos de igual tamaño para asignarlos a las diferentes casillas.

Dicotomización de variables numéricas

Supongamos que un psicólogo especializado en desarrollo midió la angustia y la habilidad social en un grupo de niños. El psicólogo después observó sus comportamientos en un grupo de juego con otros niños, concentrándose en sus reacciones agresivas. Para observar los resultados de este estudio, el investigador dividió a los niños en dos grupos según sus puntuaciones de angustia, formando un grupo de alto nivel de angustia y otro de bajo nivel de angustia; después los dividió nuevamente por la mitad según las habilidades sociales, formando un grupo con habilidades sociales altas y otro con bajas. Las combinaciones dieron como resultado cuatro grupos: alta angustia, alto nivel de habilidades sociales; alta angustia, bajo nivel de habilidades sociales, y así sucesivamente. Habiendo dividido a los niños del modo descripto, el investigador realizó un análisis de varianza 2 x 2 comparando el alto nivel contra el bajo nivel de angustia según el alto nivel contra el bajo nivel de habilidades sociales. A través del análisis, el investigador pudo observar si existía un efecto principal de la angustia sobre la agresión, un efecto principal del nivel de habilidades sociales sobre la agresión, y/o un efecto interactivo de la angustia y las habilidades sociales sobre la agresión.

Lo que debemos observar en este caso es que el investigador dividió a los niños en dos grupos según la angustia y las habilidades sociales. Analicemos primero la angustia. En el estudio, la angustia era una variable numérica cuantitativa considerada continua; sin embargo, el investigador ignoró todas las delicadas gradaciones y simplemente dividió al grupo por la mitad, formando un grupo con alto nivel de angustia y otro con bajo nivel de angustia. Como resultado, todos los participantes del grupo con alto nivel de angustia eran tratados como si tuvieran la misma puntuación, y todos los de la agrupación con bajo nivel de angustia eran tratados como si tuvieran la misma puntuación.

El tipo de división que acabamos de describir se denomina **dicotomización**, es decir, convertir la variable en una dicotomía o en dos modalidades. Dado que la dicotomización se realiza usualmente tomando a aquéllos por encima y por debajo de la mediana, el proceso descripto también se denomina **división por la mediana** de las puntuaciones. (Es menos común que los investigadores dividan por la media, pero ya sea que se utilice la mediana o la media, no se produce gran diferencia). En el ejemplo que analizamos, el investigador también dicotomizó (realizó una división por la mediana) según las habilidades sociales.

La ventaja de dicotomizar variables numéricas es que posibilita la realización de un análisis de varianza factorial, con todas las ventajas que hemos visto en este capítulo relacionadas con la eficiencia y con la posibilidad de realizar pruebas de efectos principales y de interacción en el mismo estudio. Además, la mayoría de los psicólogos están familiarizados con el análisis de varianza y comprenden rápidamente tales resultados. Muchos psicólogos están menos familiarizados con los procedimientos alternativos (basados en regresiones múltiples), que logran prácticamente el mismo resultado pero no requieren de la dicotomización.

Una desventaja importante de la dicotomización es que gran cantidad de información se pierde cuando se reduce a dos todo el rango de valores, es decir, en alto o bajo. Desde otro punto de vista, también podría decirse que la medición se hace mucho menos precisa. Uno de los resultados de este proceso es que el tamaño del efecto y la potencia de un estudio con dicotomización son mucho menores que cuando se utilizan los valores originales. Cohen (1983) calculó que la reducción de la potencia y del tamaño del efecto ¡es de entre un 20% y un 66%! Sugirió que es equivalente a "descartar uno o dos tercios de la muestra". (p. 253)

Por otro lado, muchos investigadores dicotomizan sus variables alegando que el efecto es conservador, perdiendo aumentar la probabilidad de un error Tipo II (no rechazar la hipótesis nula cuando en realidad es falsa), pero no aumenta la probabilidad del error Tipo I (rechazar la hipótesis nula cuando es verdadera). Dicho de otro modo, la dicotomización reduce la precisión de modo tal que resulta más difícil, y no más fácil, obtener resultados significativos. Por lo tanto, si un estudio descubre un resultado a pesar de la dicotomización, podemos confiar en que se trata de un efecto real. Más aún, sostienen que si el tamaño del efecto es menor de lo que podría ser sin dicotomización, al menos podemos confiar en que no hemos sobrestimado el tamaño de efecto. La dicotomización continúa siendo muy común en las investigaciones ya que facilita el cálculo y la interpretación de los resultados.

Casi todos los que han escrito acerca de esta cuestión están de acuerdo con que el análisis anterior es el correcto, pero sólo hasta cierto punto. En líneas generales, el efecto de la dicotomización de una sola variable es conservador. Por ejemplo, supongamos que el investigador sólo comparara niños con alto y bajo nivel de angustia con respecto a la agresión (sin considerar las habilidades sociales). En ese caso, el procedimiento sería conservador en el sentido de que la probabilidad de obtener significación disminuye cuando no existe efecto verdadero.

Por supuesto que una decisión como la anterior es también excesivamente "conservadora", ya que no se descubrirán los verdaderos resultados y los verdaderos tamaños del efecto serán subestimados. También sucede que existe una falta de exactitud general. El efecto de la dicotomización es conservador **en líneas generales**. Pero en algún caso en particular, la inexactitud de la dicotomización podría funcionar en favor de la hipótesis del investigador, aumentando inadecuadamente el tamaño del efecto y haciendo que una verdadera ausencia de diferencia resulte significativa en el estudio.

Más aún, Maxwell y Delaney (1993) han demostrado que cuando se dicotomizan dos variables (como en el ejemplo de la angustia y las habilidades sociales), ya no puede asumirse que el impacto sea conservador. En una cantidad de situaciones muy comunes dentro de la investigación psicológica, dicotomizar dos variables puede tener consecuencias opuestas a las de un efecto conservador, aun en líneas generales. Según Maxwell y Delaney, uno debería ser especialmente escéptico en cuanto a los resultados de estudios que utilizan un análisis de varianza de dos criterios, en el que ambas variables han sido dicotomizadas.

Existe otra desventaja de la dicotomización en el análisis factorial de varianza. En la mayoría de los casos, cuando dividimos dos variables numéricas por sus medianas, las casillas resultantes en el análisis 2 x 2 por lo general no son iguales, dando lugar a una mayor reducción de la potencia y a los problemas de susceptibilidad, y al incumplimiento de supuestos que tratamos en la sección anterior.

A pesar de las diversas dificultades mencionadas, la dicotomización (incluso la de dos variables *en un análisis de varianza* de dos criterios) es aún sorprendentemente común en la investigación psicológica. Es nuestra impresión, sin embargo, que está desapareciendo rápidamente.

RESULTADOS DEL ANÁLISIS FACTORIAL DE VARIANZA SEGÚN SE DESCRIBEN EN LAS PUBLICACIONES CIENTÍFICAS

En un análisis factorial de varianza, los resultados comúnmente se presentan con una descripción en el texto, más una tabla. El texto indica la razón F y la información que éste implica para cada efecto principal e interactivo. La tabla indica las medias de casilla y, a veces, también las medias marginales. Si existe un efecto interactivo, en lugar de (o además de) una tabla puede haber un gráfico. Por Ejemplo, Lambert et al. describieron de la siguiente manera el resultado que utilizamos como ejemplo:

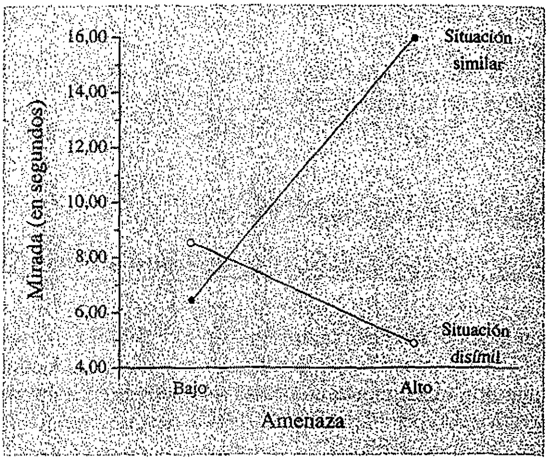
> El análisis de las intenciones de los participantes de contratar al objetivo reveló sólo un efecto significativo, la interacción predicha entre el estado de ánimo y el tipo de empleo, $F(1, 57) = 11{,}46$, $p < 0{,}001$. La información pertinente a esta interacción aparece en la figura [13-2] (p. 1011).

Hasta hace poco, era más común que los investigadores mostraran los efectos interactivos utilizando una especie de gráfico lineal. Analicemos un ejemplo: Gump y Kulik (1997) realizaron un estudio para probar una teoría acerca de condiciones que promovían las relaciones interpersonales. Específicamente, predijeron que es más probable que uno se relacione con otra persona cuando uno está bajo amenaza y cuando la otra persona está enfrentando la misma amenaza. Como parte del estudio, se asignaron participantes al azar a esperar o no un procedimiento experimental doloroso. La anterior era la manipulación en cuanto a la amenaza (alta contra baja). En todas las condiciones había otro participante en la habitación. A la mitad de los participantes se les dijo que ese otro participante estaba en el mismo experimento y, por lo tanto, enfrentaba una situación similar de amenaza o de no amenaza; la otra mitad suponía que el otro participante formaba parte de un experimento totalmente diferente y, por lo tanto, no esperaba la misma situación de amenaza o no amenaza. Esta era la manipulación en cuanto a la similitud de la situación (similar contra disímil). A medida que se informaba al participante acerca de la amenaza a través de un auricular, los experimentadores observaron cuánto tiempo pasaba cada participante mirando a su compañero.

> En este análisis surgió una interacción significativa entre la amenaza hacia el participante y la similitud de la situación, $F(1, 77) = 5{,}57$, $p = 0{,}02$. Ningún otro efecto resultó significativo. Como se observa en la figura [13-11], queda claro que un alto nivel de amenaza produjo que se mirara más a un compañero al que se creía en la misma situación, en comparación con aquél al que se creía en una situación diferente, mientras que esa diferenciación no ocurrió en el caso de participantes bajo una amenaza de bajo nivel.

El tipo de gráfico que presentaron Gump y Kulik (figura 13-11) tiene la ventaja de hacer que el patrón de la interacción sea muy claro, aun comparado con el usual gráfico de barras. La razón por la que los gráficos de líneas como el del ejemplo se han vuelto menos comunes en los últimos años, es que son levemente engañosos, en el sentido de que la línea implica que existe un efecto continuo. Por ejemplo, en el estudio de Gump y Kulik había una condición de alto nivel de amenaza y otra de bajo nivel de amenaza. Las líneas dan la impresión de que el patrón de cada condi-

Figura 13-11.
Estudio 1: efecto de la amenaza y la similitud de situación en el tiempo transcurrido mirando a un compañero. [**Fuente**: Gump, B. B. & Kulik, J. A. (1997), fig. 1. "Estrés, relación y contagio emocional". *Revista Científica de Psicología Social y de la Personalidad (Journal of Personality and Social Psychology)*, 72, 305-319. Copyright, 1997, por la Asociación Americana de Psicología. Reimpreso con autorización].



ción de similitud es continuo de bajo a alto nivel de amenaza. Supongamos que los investigadores hubieran incluido una condición intermedia de amenaza. Es posible que el resultado para esa condición no estuviera en absoluto en el lugar en el que la línea del gráfico en cuestión indica que debería estar. Los gráficos de barra, por el contrario, no reflejan nada en cuanto a los niveles intermedios, simplemente muestran los resultados de cada nivel de la variable probada.

Resumen

En un diseño de investigación factorial los participantes son divididos en grupos según las combinaciones de las variables cuyos efectos están siendo analizados. A través de los diseños factoriales podemos analizar los efectos de dos (o más) variables sin necesidad de convocar el doble de participantes. Además, estos diseños hacen posible el análisis de efectos interactivos, es decir, los efectos de las combinaciones de las dos variables. Específicamente, un efecto interactivo ocurre cuando el efecto de una variable depende del nivel de la otra variable. Un efecto principal es el efecto promedio general de una variable, ignorando el efecto de la otra variable. Los efectos principales e interactivos pueden describirse verbal, numérica y gráficamente.

Los cálculos de un análisis de varianza de dos criterios siguen el método del modelo estructural. La estimación intragrupal de varianza poblacional es, en realidad, una estimación de varianza poblacional a partir del interior de las casillas. Se basa en los desvíos de cada registro con respecto a la media de su casilla. Existen tres diferentes estimaciones intergrupales de varianza: una para las diferencias de la dispersión de la variable a través de las filas, otra para las diferencias de la dispersión de la variable a través de las columnas y una para la interacción de las variables de fila y de columna. El efecto de las filas se basa en los desvíos entre las medias de las filas y la gran media, y el efecto principal de las columnas se basa en los desvíos entre las medias de las columnas y la gran media. El efecto interactivo se basa en el desvío restante entre las observaciones y la gran media después de restar todos los otros desvíos con respecto a la gran media (desvíos de las medias de las casillas, las medias de filas y las medias de columnas). Para obtener las estimaciones reales de varianza poblacional, esos distintos desvíos (interno, filas, columnas e interactivo) se elevan al cuadrado, se suman y se dividen por sus grados de libertad. Las razones F para los efectos de fila, columna e interactivo se calculan dividiendo las estimaciones de varianza poblacional, correspondientes a cada uno de ellos, por la estimación intragrupal de casilla de la varianza poblacional.

En un análisis factorial de varianza se calcula el tamaño del efecto y la potencia separadamente para cada efecto principal e interactivo. El indicador más útil del tamaño del efecto es la proporción de varianza explicada, R^2 (también denominada eta^2). En un análisis de varianza de dos criterios, se calcula R^2 para cualquier efecto principal o interactivo en particular de la siguiente forma: se divide la suma de los cuadrados correspondientes a ese efecto en particular por la parte de la suma total de cuadrados que queda después de restarle la suma de los cuadrados correspondientes a los otros dos efectos.

El análisis factorial de varianza puede extenderse a diseños de más de dos criterios e incluso puede utilizarse para manejar estudios de medidas repetidas.

Existen dos antiguas controversias con respecto al análisis factorial de varianza. Una se basa en la forma de manejar situaciones con tamaños desiguales de casilla. El método de cuadrados mínimos se considera usualmente el mejor, pero la solución óptima es trabajar con casillas del mismo tamaño. La otra controversia se basa en la conveniencia de realizar una dicotomización de variables continuas para realizar un análisis de varianza. El procedimiento de dicotomización cada vez está siendo menos común; generalmente se considera mejor utilizar procedimientos más avanzados que conservan todos los valores de cada variable.

Los resultados del análisis factorial de varianza incluyen con frecuencia descripciones gráficas de los resultados, particularmente cuando el efecto interactivo es significativo. Por lo general se utilizan gráficos de barra, pero, a veces, se incluyen gráficos de líneas.

Términos clave

- Casilla.
- Media de casilla.
- Dicotomización.
- Análisis factorial de varianza.
- Diseño factorial de investigación.
- Efecto interactivo.
- Análisis de varianza de cuadrados mínimos.
- Efecto principal.
- Medias marginales.
- División por la mediana.
- Análisis de varianza de un criterio.
- Análisis de varianza de medidas repetidas.
- Diseño factorial de tres criterios.
- Análisis de varianza de dos criterios.
- Diseño de investigación factorial de dos criterios.

Ejercicios

Los ejercicios implican la realización de cálculos (con la ayuda de una calculadora). La mayoría de los problemas estadísticos reales se resuelven por computadora, pero aunque exista la posibilidad de utilizarla, es conveniente realizar estos ejercicios manualmente para incorporar el método de trabajo.

Para adquirir práctica en la utilización de una computadora, para resolver problemas estadísticos, se puede utilizar la sección de computación de cada capítulo, publicada en la *Guía de estudio y libro de tareas de computación para el alumno [Student's Study Guide and Computer Workbook]* que acompaña este libro.

Todos los datos de esta sección son ficticios (a menos que se especifique lo contrario)

Las respuestas a los ejercicios de la serie I se encuentran al final del libro.

SERIE I

1. Cada una de las siguientes es una tabla de medias que muestra los resultados de un estudio con diseño factorial. Suponiendo que cualquier diferencia es estadísticamente significativa, para cada tabla a) realice dos gráficos de barra que muestren los resultados (en un gráfico agrupe las barras según una variable y en el otro gráfico agrupe las barras según la otra variable); b) indique qué efectos se encontraron (principales e interactivos), si los hay, y c) describa el significado del patrón de medias (es decir, la existencia de cualquier efecto principal o interactivo, o su ausencia) con palabras.

i) Variable medida: ingreso (miles de dólares).

		Edad	
		Joven	*Mayor*
Clase	*Inferior*	20	35
	Superior	25	100

ii) Variable medida: promedio de calificaciones.

		Especialidad	
		Ciencia	*Arte*
Facultad	*Comunidad*	2,1	2,8
	Artes Liberales	2,8	2,1

iii) Variable medida: días de enfermedad por mes.

		Sexo	
		Femenino	*Masculino*
Grupo	*Pasantes*	2,0	2,5
	Supervisores	3,1	3,6

iv) Variable medida: calificación de la calidad del restaurante (de 1 a 10).

		Ciudad		
		New York	*Chicago*	*Vancouver*
Costo	*Caro*	9	5	7
	Moderado	6	4	6
	Barato	4	3	5

2. Un psicólogo especializado en deportes realizó un estudio acerca del efecto de un programa de motivación en las lesiones entre jugadores de tres deportes diferentes. El cuadro que sigue a continuación muestra el diseño. Para cada uno de los siguientes posibles patrones de resultados, cree una serie de medias de casilla, calcule las medias marginales y realice un gráfico de barras de los resultados: a) efecto principal del tipo de deporte y ningún otro efecto principal o interactivo; b) efecto principal del programa o la ausencia de programa y ningún otro efecto principal o interactivo; c) ambos efectos principales sin interacción; d) efecto del programa o la ausencia de programa y una interacción, pero sin efecto principal del tipo de deporte; e) ambos efectos principales y una interacción.

Variable medida: Cantidad de lesiones por persona durante 10 semanas.

	Deporte		
	Baseball	*Fútbol*	*Basket*
Con programa de motivación			
Sin programa de motivación			

3. a) ¿Cuál sería la potencia de cada uno de los efectos en el estudio del ejercicio 2 si el investigador tiene 40 participantes por casilla utilizando un nivel de 0,05, y suponiendo tamaños del efecto medianos? b) ¿Cuántos participantes en total serían necesarios para un 80% de potencia si el investigador espera que todos los tamaños del efecto sean grandes?

4. Un psicólogo realiza un estudio para comparar la efectividad relativa de tres tipos diferentes de terapia en pacientes con diferentes diagnósticos. Pacientes con dos tipos de diagnósticos fueron asignados al azar a uno de los tres tipos de terapia. Había dos pacientes por casilla. Sobre la base de los resultados que aparecen abajo a) realice el análisis de varianza; b) realice una tabla de medias de casilla y marginales; c) calcule los tres tamaños de efecto, y d) describa los resultados con palabras (*indique qué efectos son significativos y*, en base a ellos, cómo comprender el patrón de medias de casilla). Utilice el nivel 0,05.

	Terapia A	Terapia B	Terapia C
Diagnóstico I	6	3	2
	2	1	4
Diagnóstico II	11	7	8
	9	9	10

5. Un psicólogo que analiza el sistema judicial realiza un estudio acerca del efecto de la simpatía y del nerviosismo del acusado en la predisposición a condenarlo. Cada participante leyó la misma transcripción tomada de un juicio real en el que la culpabilidad o inocencia de un acusado de sexo masculino era bastante ambigua. Todos los participantes vieron también un breve video que supuestamente mostraba al acusado en el estrado. Sin embargo, la actuación del protagonista del video era diferente para los distintos participantes, incluyendo las cuatro posibilidades de simpatía contra la falta de ella y el nerviosismo contra la ausencia del mismo. Después de observar la cinta, los participantes calificaron la posibilidad de que el acusado fuera inocente (en una escala de 1, muy improbable, a 10, muy probable). Los resultados de los primeros 12 participantes del estudio fueron los siguientes:

	Simpatía	Falta de simpatía
Nerviosismo	7	3
	8	4
	6	2
Ausencia de	3	7
nerviosismo	3	5
	3	9

a) Realice el análisis de varianza. b) Construya una tabla de medias de casilla y marginales. c) Calcule los tamaños del efecto. d) Explique los resultados y la forma en que llegó a los mismos a alguien que está familiarizado con el análisis de varianza de un criterio (incluyendo el método del modelo estructural) pero no con el análisis factorial de varianza.

6. Kunda y Oleson (1997) realizaron un estudio acerca de estereotipos, que se concentraba en el efecto de información contraria al estereotipo. Por información contraria definían el hecho de conocer características de determinada persona, las cuales eran contrarias a lo que se esperaría del estereotipo. Específicamente, analizaron la posibilidad de que la información extremadamente contraria pueda tener un efecto *boomerang*, es decir, reforzar el estereotipo.

Se preseleccionaron los participantes para el estudio sobre la base de respuestas a un cuestionario que incluía una pregunta en la que calificaban a los agentes RRPP (Relaciones Públicas) en cuanto a su grado típico de extroversión. La mayoría de las personas consideran que los agentes RRPP son extrovertidos. Sin embargo, los investigadores seleccionaron un grupo de participantes, los partici- pantes de "estereotipo extremo", que habían clasificado a los agentes RRPP como extremadamente extrovertidos. El otro grupo, denominado de "estereotipo moderado", había clasificado a los agentes RRPP como sólo moderadamente introvertidos. Durante el estudio real, a algunos participantes se les dio una descripción de un determinado agente RRPP que era muy introvertido y, por lo tanto, desviado en extremo de las expectativas usuales que indican que los agentes RRPP son extrovertidos. Era la condición de desvío extremo. A los otros participantes no se les dio ninguna descripción en especial, era la condición de control. Mas tarde, se preguntó a todos los participantes qué pensaban con respecto a los agentes RRPP.

Kunda y Oleson (1997) informaron los resultados de la siguiente manera:

> Un ANOVA de 2 (estereotipo previo) x 2 (condición) arrojó una interacción significativa, $F(1, 42) = 5{,}69$, $p < 0{,}05$, indicando que el impacto del objetivo en los estereotipos de los participantes dependía de sus estereotipos previos. Como se observa en la figura [13-12], los participantes de estereotipo extremo expuestos al objetivo altamente introvertido llegaron a considerar a los agentes RRPP como aún más extrovertidos de lo que lo hicieron los controles de extremos...[a] efecto *boomerang* [...] Se observó un patrón diferente para los participantes de estereotipo moderado. Sus estereotipos no fueron afectados por la exposición al mismo objetivo [...] El ANOVA también reveló

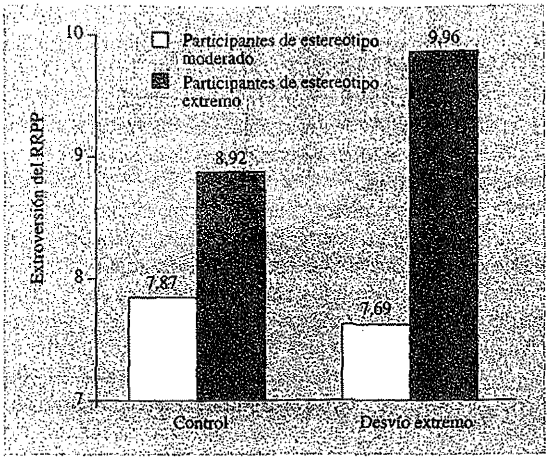


Figura 13-12.
Calificaciones medias de la extroversión de agentes RRPP realizada por participantes con estereotipos previos, moderados o extremos, *que fueron expuestos a un* agente RRPP extremadamente introvertido o a ningún objetivo (controles). Los números más altos indican mayor extroversión de los RRPP. [**Fuente**: Kunda, Z., & Oleson, K. C. (1997), figura 4. "Cuando las excepciones confirman la regla: cómo el extremismo de una desviación determina el impacto de los ejemplos desviados en los estereotipos". *Revista Científica de Psicología Social y de Personalidad [Journal of Personality and Social Psychology]*, 72, 965-979. Copyright, 1997, por la Asociación Americana de Psicología. Reimpreso con autorización.

> un gran efecto de los estereotipos previos, F(1, 42<9 = 38,94, $p < 0,0001$, indicando, lo cual no es sorprendente, que los participantes de estereotipo extremo continuaron considerando a los agentes RRPP como más extrovertidos de lo que lo hicieron los participantes de estereotipo moderado. También hubo un efecto marginal de la condición, $F(1, 42) = 2,89$, $p < 0,10$, que se debía claramente en su totalidad a los participantes de estereotipo extremo. (p. 974).

Describa brevemente el significado de estos resultados a una persona que nunca ha asistido a un curso de estadística. (Evite entrar en detalles de cálculo; sólo incluya la lógica básica del patrón de medias, los resultados significativos, los tamaños del efecto y los aspectos relacionados con la interpretación de resultados no significativos).

SERIE II

1. Cada una de las siguientes es una tabla de medias que muestra los resultados de un estudio con diseño factorial. Suponiendo que cualquier diferencia es estadísticamente significativa, para cada tabla a) realice dos gráficos de barra mostrando los resultados (en un gráfico agrupe las barras según una variable y en el otro gráfico agrupe las barras según la otra variable); b) indique qué efectos se encuentran (principales e interactivos), si existen; c) describa el significado del patrón de medias (es decir, cualquier efecto principal o interactivo o la ausencia de los mismos) con palabras.

i) Variable medida: intensidad de atención.

Tipo de asistente a espectáculos de ballet	**Programa**	
	El cascanueces	*Moderno*
Regular	20	15
A veces	15	15
Neófito	10	5

ii) Variable medida: nivel de aprobación del presidente de EEUU.

	Región			
	Oeste	*Este*	*Los estados centrales*	*Sur*
Media	70	45	55	50
Baja	50	25	35	30

iii) Variable medida: satisfacción con respecto a la educación.

Tiempo transcurrido después de obtener el BA	Sexo	
	Femenino	*Masculino*
1 mes	3	3
1 año	4	4
5 años	9	9

iv) Variable medida: nivel de envidia del éxito de otra persona.

Relación con el otro	Nivel de éxito	
	Grande	*Pequeño*
Amigo	8	5
Extraño	1	4

2. En este estudio se instruyó a participantes de habla inglesa para que intentaran leer durante media hora un párrafo escrito en uno de tres idiomas que desconocían. Leyeron el párrafo después de que se les dijera la idea principal de todo el párrafo o sólo la idea principal de la primera oración, o bien después de que no se les dijera nada sobre el significado del párrafo. Se les dio la traducción de algunas palabras. Después los investigadores midieron cuántas de las otras palabras pudieron traducir correctamente. El cuadro que sigue a continuación muestra el diseño. Para cada uno de los siguientes posibles patrones de resultado, cree una serie de medias de casilla, calcule las medias marginales y realice un gráfico de barras de los resultados: a) efecto principal del idioma y ningún otro efecto principal o interactivo; b) efecto principal del conocimiento del significado y ningún otro efecto principal o interactivo; c) ambos efectos principales pero sin interacción; d) efecto principal del idioma y una interacción, pero sin efecto principal del conocimiento del significado; e) ambos efectos principales y una interacción.

Conocimiento del significado	Idioma		
	Holandés	*Rumano*	*Sueco*
Párrafo			
Oración			
Ninguno			

3. En determinado colegio secundario, se probaron tres tipos de programas de enseñanza de inglés, historia y matemática a través de videos. Después, los investigadores midieron el nivel de aprendizaje. Había dos alumnos por casilla. Sobre la base de los resultados que aparecen abajo, a) realice una tabla de medias de casilla y marginales y trace un gráfico de barras de las mismas; b) realice los cinco pasos de la prueba de hipótesis (utilice el nivel 0,05 de significación); c) calcule los tres tamaños del efecto, y d) describa los resultados con palabras (indique qué efectos son significativos y, sobre la base de ellos, cómo interpreta el patrón de medias de casillas).

	Inglés	**Historia**	**Matemática**
Programa Tipo A	3	15	2
	3	14	3
Programa Tipo B	6	18	6
	8	10	5
Programa Tipo C	1	13	2
	3	4	0

4. Para cada una de las siguientes series de información realice un análisis de varianza, incluyendo una tabla de medias de casilla y marginales y un gráfico de barras de las medias de casilla.

(i) **Condición Experimental**

	A	*B*
Grupo 1	0	3
	1	2
	1	3
Grupo 2	3	0
	2	1
	3	1

(ii) **Condición Experimental**

	A	*B*
Grupo 1	0	0
	1	1
	1	1
Grupo 2	3	3
	2	2
	3	3

(iii) **Condición Experimental**

	A	B
Grupo 1	0	3
	1	2
	1	3
Grupo 2	0	3
	1	2
	1	3

5. Desmarais y Curtis (1997), dos psicólogos sociales canadienses, realizaron un estudio relacionado con la forma en que mujeres y hombres evalúan su propio valor como empleados. Estudios previos habían demostrado que cuando se les daba la posibilidad de decidir cuánto deberían ganar por una tarea experimental, las mujeres usualmente se establecían una paga menor. Los investigadores esperaban que este patrón fuera afectado por el monto en particular que mujeres y hombres habían recibido como pago por trabajo real en el pasado reciente. Realizaron el procedimiento experimental habitual en el que se le pidió a estudiantes mujeres y hombres que se asignaran un sueldo, pero también se preguntó a los participantes cuánto habían ganado el verano anterior. Desmarais y Curtis informaron sus resultados de la siguiente manera:

> Se analizaron los pagos auto-asignados por los participantes con un ANOVA 2 x 3 (sexo de los participantes x ingreso percibido recientemente). En contra de la predicción, en cuanto a que el ingreso reciente influiría en el pago asignado por ellos mismos, la información reveló que no existía diferencia de adjudicación de pago causada por el ingreso percibido, $F(2, 66) = 1{,}99$, *ns* (*véase* [tabla 13-14]). Además, no hubo interacción significativa entre sexo y el ingreso percibido, $F(2, 66) = 0{,}61$, *ns*. En concordancia con investigaciones previas sobre percepción de ingreso merecido, los hombres se pagan a sí mismos montos significativamente más altos ($M = \$3{,}99$) que las mujeres ($M = \$2{,}74$), $F(1, 66) = 5{,}86$, $p < 0{,}02$. (p. 143).

Describa brevemente el significado de estos resultados a una persona que nunca ha asistido a un curso de estadística. (Evite entrar en detalles de cálculo; sólo incluya la lógica básica del patrón de medias, los resultados significativos, los tamaños de efecto y las cuestiones relacionadas con la interpretación de los resultados no significativos).

Tabla 13-14.
Media de pago auto-asignado por una tarea experimental, dividido por categorías según el sexo y el ingreso percibido el verano anterior.

	Ingreso del verano enterior					
	Bajo		*Mediano*		*Alto*	
Sexo	*M*	*SD*	*M*	*SD*	*M*	*SD*
Hombres	5,03	1,71	3,17	3,00	3,77	2,77
Mujeres	3,13	1,68	2,65	1,89	2,44	1,65

Nota: Los estudiantes de la categoría de bajos ingresos ganaron menos de \$6,00/hr; los estudiantes de la categoría de medianos ingresos ganaron entre \$7,50 y \$8,50/hr; los alumnos de la categoría de altos ingresos ganaron más de \$10,00 la hora. Para cada casilla, $n = 12$. Sólo fue significativo el efecto principal del sexo, $p < 0{,}02$.

Fuente: Desmarais, S. & Curtis, J. (1997), tab. 1. "Sexo y percepción del ingreso merecido: prueba de los efectos del ingreso percibido". *Revista científica de Psicología Social y de la Personalidad [Journal of Personality and Social Psychology]*, 72, 141-150. Copyright, 1997, por la Asociación Americana de Psicología. Reimpreso con autorización.

Apéndice I del capítulo: fórmulas optativas de cálculo para el análisis de varianza de dos criterios

Esta sección proporciona las fórmulas de cálculo para un análisis de varianza de dos criterios, que no requieren el cálculo de desvíos para cada individuo. Al igual que las otras fórmulas presentadas en este capítulo, sólo se aplican cuando hay igual cantidad de participantes por casilla.

La fórmula de cálculo para la suma total de desvíos cuadráticos es igual a la fórmula 12-6 (y la misma que para SC en la fórmula 12-9).

$$SC_{\text{Total}} = \Sigma X^2 - \frac{(\Sigma X)^2}{N} \tag{13-26}$$

En la fórmula anterior, ΣX^2 es la suma de los cuadrados de todas las observaciones; $(\Sigma X)^2$ es el cuadrado de la suma de todas las observaciones, y N es la cantidad total de observaciones.

La fórmula para el efecto intergrupal general también es igual a la anterior (fórmula 12-7), excepto que los subíndices ahora se refieren a casillas en lugar de grupos.

$$SC_{\text{Entre}} = \frac{(\Sigma X_1)^2}{n} + \frac{(\Sigma X_2)^2}{n} + \ldots + \frac{(\Sigma X_{\text{Último}})^2}{n} - \frac{(\Sigma X)^2}{N} \tag{13-27}$$

En esta fórmula, $(\Sigma X_1)^2 + (\Sigma X_2)^2, \ldots, + (\Sigma X_{\text{Última}})^2$ son los cuadrados de las sumas de las observaciones de cada casilla; n es la cantidad de participantes de cada casilla.

La suma de los cuadrados intragrupales es el total menos el intergrupal:

$$SC_{\text{Dentro}} = SC_{\text{Total}} - SC_{\text{Entre}} \tag{13-28}$$

La fórmula para la suma de los cuadrados de filas es una versión modificada de la fórmula general de los intergrupales (13-27):

$$SC_{\text{Filas}} = \frac{(\Sigma X_{\text{Filas}_1})^2}{n_{\text{Fila}}} + \frac{(\Sigma X_{\text{Filas}_2})^2}{n_{\text{Fila}}} + \ldots + \frac{(\Sigma X_{\text{Fila Última}})^2}{n_{\text{Fila}}} - \frac{(\Sigma X)^2}{N} \tag{13-29}$$

En la fórmula anterior: $(\Sigma X_{\text{Fila}_1})^2, (\Sigma X_{\text{Fila}_2})^2 \ldots, (\Sigma X_{\text{Fila}_{\text{Última}}})^2$ son los cuadrados de las sumas de todas las observaciones de cada fila; n_{Fila} es la cantidad de participantes en cada fila.

La fórmula general intergrupal para columnas sigue el mismo principio:

$$SC_{\text{Columnas}} = \frac{(\Sigma X_{\text{Columna}_1})^2}{n_{\text{Columna}}} + \frac{(\Sigma X_{\text{Columna}_2})^2}{n_{\text{Columna}}} + \ldots + \frac{(\Sigma X_{\text{Columna última}})^2}{n_{\text{Columna}}} - \frac{(\Sigma X)^2}{N} \tag{13-30}$$

Tabla 13-15.
Cálculo de sumas de cuadrados para un análisis de varianza de dos criterios basado en Blanchard et al. (1991), utilizando fórmulas de cálculo. (Datos ficticios).

	Modalidad de respuesta pública		Modalidad de respuesta privada		Filas	
	X	X^2	X	X^2		
Influencia antirracista						
	25	625	19	361		
	20	400	24	576		
	23	529	21	441	**Influencia antirracista**	
	24	576	20	400	ΣX	ΣX^2
Casilla Σ:	92	2.130	84	1.778	176	3.908
Ausencia de influencia						
	22	484	24	576		
	19	361	18	324		
	22	484	22	484	**Ausencia de influencia**	
	21	441	20	400	ΣX	ΣX^2
Casilla Σ:	84	1.770	84	1.784	168	3.554
Ausencia no antirracista :						
	16	256	18	324		
	19	361	21	441		
	13	169	16	256	**Ausencia no antirracista**	
	16	256	17	289	ΣX	ΣX^2
Casilla Σ:	64	1.042	72	1,310	136	2.352
Columna Σ:	240	4.942	240	4.872		

Total $\Sigma X = 480$
Total $\Sigma X^2 = 9.814$

$$SC_{\text{Total}} = \Sigma X^2 - \frac{(\Sigma X)^2}{N} = 9.814 - \frac{480^2}{24} = 9.814 - \frac{230.400}{24} = 9.814 - 9.600 = 214$$

$$SC_{\text{Entre}} = \frac{92^2}{4} + \frac{84^2}{4} + \frac{84^2}{4} + \frac{84^2}{4} + \frac{64^2}{4} + \frac{72^2}{4} - \frac{480^2}{24}$$

$$= \frac{8.464}{4} + \frac{7.056}{4} + \frac{7.056}{4} + \frac{7.056}{4} + \frac{4.096}{4} + \frac{5.184}{4} - \frac{230.400}{24}$$

$$= 2.116 + 1.764 + 1.764 + 1.764 + 1.024 + 1.296 - 9.600 = 128$$

$$SC_{\text{Dentro}} = SC_{\text{Total}} - SC_{\text{Entre}} = 214 - 128 = 86$$

$$SC_{\text{Filas}} = \frac{(\Sigma X_{\text{Fila1}})^2}{n_{\text{Filas}}} + \frac{(\Sigma X_{\text{Fila2}})^2}{n_{\text{Filas}}} + \ldots + \frac{(\Sigma X_{\text{Fila última}})^2}{n_{\text{Filas}}} - \frac{(\Sigma X)^2}{N} = \frac{176^2}{8} + \frac{168^2}{8} + \frac{136^2}{8} - \frac{480^2}{24}$$

$$= \frac{30.976}{8} + \frac{28.224}{8} + \frac{18.496}{8} - \frac{230.400}{24} = 3.872 + 3.528 + 2.312 - 9.600 = 112$$

$$SC_{\text{Columnas}} = \frac{(\Sigma X_{\text{Columna 1}})^2}{n_{\text{Columna}}} + \frac{(\Sigma X_{\text{Columna 2}})^2}{n_{\text{Columna}}} + \ldots + \frac{(\Sigma X_{\text{Columna última}})^2}{n_{\text{Columna}}} - \frac{(\Sigma X)^2}{N}$$

$$= \frac{240^2}{12} + \frac{240^2}{12} - \frac{480^2}{24} = \frac{57.600}{12} + \frac{57.600}{12} - \frac{230.400}{24}$$

$$= 4.800 + 4.800 - 9.600 = 0$$

$$SC_{\text{Interacción}} = SC_{\text{Entre}} - SC_{\text{Filas}} - SC_{\text{Columnas}} = 128 - 112 - 0 = 16$$

En la fórmula anterior: $(\Sigma X_{\text{Columna}_1})^2$, $(\Sigma X_{\text{Columna}_2})^2$,..., $(\Sigma X_{\text{ColumnaÚltima}})^2$ son los cuadrados de las sumas de todas las observaciones de cada columna; n_{Columna} es la cantidad de participantes de cada columna.

Finalmente, la suma de cuadrados para la interacción se calcula sobre la base de lo que queda después de restar las sumas de cuadrados de filas y columnas a la suma general de cuadrados intergrupales:

$$SC_{\text{Interacción}} = SC_{\text{Entre}} - SC_{\text{Filas}} - SC_{\text{Columnas}} \tag{13-31}$$

La tabla 13-15 muestra los cálculos utilizando las fórmulas anteriores para el cálculo de las sumas de los cuadrados del ejemplo que utiliza datos ficticios basados en el estudio de Blanchard et. al. (1991). Compare esos cálculos con los de la tabla 13-9, que aplica las fórmulas de definición.

Apéndice II del capítulo: análisis de varianza de medidas repetidas de un criterio

Podemos realizar un análisis de varianza de medidas repetidas de un criterio utilizando los procedimientos para el análisis de varianza de dos criterios, modificado en un aspecto crucial. Igual que con un análisis de varianza de dos criterios común, tratamos los grupos (condiciones) como factores, usualmente como columnas. La modificación radica en las filas; en lugar de ser otros factores, son participantes. Es decir, hay un participante por fila y cada participante tiene un valor en cada columna. Por lo tanto, las casillas tienen sólo una observación cada una.

La suma de los cuadrados de las columnas (el factor grupo de medidas repetidas) será la misma que calculamos comúnmente. Será el numerador en el análisis de varianza. Sin embargo, no podemos utilizar la variación dentro de la casilla como denominador. ¡Precisamente porque no existe variación dentro de la casilla! Hay una sóla observación por casilla, y no puede existir ninguna variación con una sola observación. En su lugar, el denominador se basa en la suma de cuadrados de la interacción. La suma de valores de la interacción también se calcula de la forma acostumbrada. Se basa en el desvío que queda después de restar los otros desvíos. En este caso, son los desvíos que quedan después de restar los desvíos intergupales de las columnas (el factor grupo) e intergrupales de las filas (de los participantes). No restamos los desvíos intragrupales de las casillas, ya que con una observación por casilla no hay nada que restar en ese caso. El impacto de todo este procedimiento es que el término por el cual realizamos la división, es decir, el denominador, no incluye la variación entre participantes (la variación de fila). El denominador es menor y, por lo tanto, es probable que el F sea mayor.

En resumen, el procedimiento para calcular un análisis de varianza de medidas repetidas es el siguiente:

1. Colocar los nombres de los grupos de observaciones en la parte superior (el factor de medidas repetidas) y destinar una fila para cada participante.

2. Calcular de la forma acostumbrada las sumas de los cuadrados correspondientes al total (desvíos cuadráticos de cada observación con respecto a la gran media), a las columnas (desvíos cuadráticos de la media de la columna de cada registro con respecto a la gran media) y a las filas (desvíos cuadráticos de la media de la fila de cada observación con respecto a la gran media). Cabe destacar que al calcular la suma de los cuadrados de las filas, las medias de las filas son iguales al valor medio de cada participante.

3. Calcular la suma de los cuadrados correspondiente a la interacción, calculando el desvío de cada observación como su desvío con respecto a la gran media menos el desvío de la media de su columna con respecto a la gran media y menos el desvío de la media de su fila con respecto a la gran media.

4. Calcular de la forma acostumbrada los grados de libertad correspondientes a las columnas (la cantidad de columnas menos 1), a las filas (cantidad de filas menos 1) y a la interacción (cantidad de casillas menos los grados de libertad de las filas y las columnas, menos 1).

5. Calcular de la forma acostumbrada los cuadrados medios (las estimaciones de varianza poblacional) correspondientes a las columnas y a la interacción. En el caso de las columnas, implica la suma de cuadrados de columnas dividida por los grados de libertad de las columnas; en el caso de la interacción, implica la suma de cuadrados de interacción dividida por los grados de libertad de la interacción. Cabe destacar que no calculamos la estimación de varianza poblacional basada en las filas debido a que sería la varianza entre participantes, que no es de nuestro interés.

6. Calcular la razón F para el efecto de medidas repetidas. Dividir el cuadrado medio correspondiente a las columnas por el cuadrado medio correspondiente a la interacción.

7. Comparar el F con el punto F de corte basándose en los grados de libertad del numerador (columnas) y del denominador (interacción) adecuados.

La tabla 13-16 indica las observaciones, cálculos y una tabla de análisis de varianza, para un análisis de varianza de medidas repetidas correspondiente a un ejemplo ficticio. El ejemplo utilizado es un estudio de errores al reconocer una sílaba inserta en uno de tres diferentes tipos de palabras: una palabra familiar, una palabra no familiar, o un sonido que no es una palabra. (Supondremos que cada tipo de palabra es presentado 30 veces de forma tal que haya un mínimo de 0 errores y un máximo de 30). Los participantes son cuatro.

Tabla 13-16.
Análisis de varianza de medidas repetidas de un criterio sobre un estudio de errores de reconocimiento por parte de cuatro participantes al ser expuestos cada uno a sílabas objetivo insertas en palabras familiares, no familiares y sonidos que no forman palabras. (Datos ficticios).

Medida: cantidad de errores

	Tipo de palabra				
Participante	*Palabra familiar*	*Palabra no familiar*	*Sonido no palabra*	*Fila* Σ	*Fila* M
A	9	3	0	12	4
B	6	2	1	9	3
C	11	6	4	21	7
D	10	5	3	18	6
Σ	36	16	8		
M	9	4	2		

$GM = 5$

Desvíos cuadráticos con respecto a la gran media

Participante	**Palabra familiar** X	Col	fil	Int	**Palabra no familiar** X	Col	fil	Int	**Sonido no palabra** X	Col	fil	Int
A	16	16	1	1	4	1	1	0	25	9	1	1
B	1	16	4	1	9	1	4	0	16	9	4	1
C	36	16	4	0	1	1	4	0	1	9	4	0
D	25	16	1	0	0	1	1	0	4	9	1	0
Σ	78	64	10	2	14	4	10	0	46	36	10	2

$SC_{\text{Total}} = 78 + 14 + 46 = 138$
$SC_{\text{Columnas}} = 64 + 4 + 36 = 104$
$SC_{\text{Filas}} = 10 + 10 + 10 = 30$
$SC_{\text{Interacción}} = 2 + 0 + 2 = 4$

Control: $SC_{\text{Total}} = SC_{\text{Columnas}} + SC_{\text{Filas}} + SC_{\text{Interacción}} = 104 + 30 + 4 = 138$

$gl_{\text{Total}} = 12 - 1 = 11$
$gl_{\text{Columnas}} = 3 - 1 = 2$
$gl_{\text{Filas}} = 4 - 1 = 3$
$gl_{\text{Interacción}} = 11 - 2 - 3 = 6$

Control: $gl_{\text{Total}} = gl_{\text{Columnas}} + gl_{\text{Filas}} + gl_{\text{Interacción}} = 2 + 3 + 6 = 11$

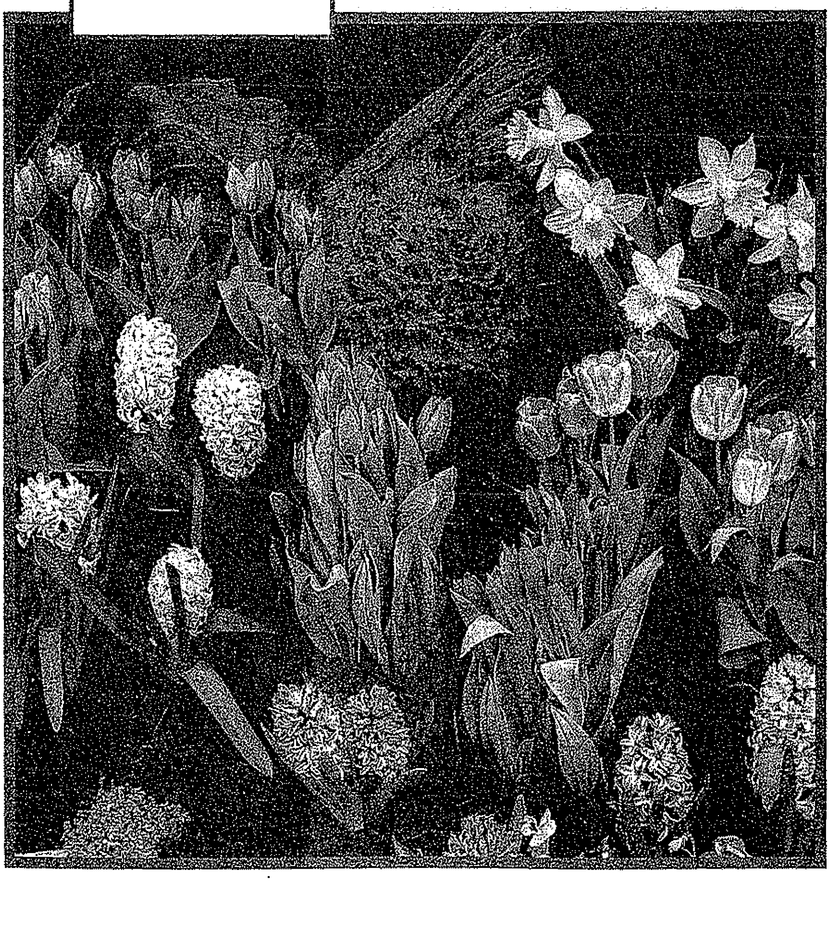
Punto de corte F para el efecto de medidas repetidas (columnas) (gl 2, 6; $p < 0{,}05$) : 5,14

Tabla del análisis de varianza

Fuente	SC	gl	CM	F
Condiciones intergrupales (columnas)	104	2	52	77,6
Participantes (filas)	30	3		
Error (Interacción)	4	6	0,67	
Total	138	11		

Conclusión: Se rechaza la hipótesis nula.

14 Pruebas chi-cuadrado



Descripción del capítulo

- El estadístico chi-cuadrado y la prueba chi-cuadrado de bondad de ajuste.
- Prueba chi-cuadrado de independencia.
- Supuestos de las pruebas chi-cuadrado.
- Tamaño del efecto y potencia en la prueba chi-cuadrado de independencia.
- Controversias y limitaciones.
- Pruebas chi-cuadrado según se describen en publicaciones científicas.
- Resumen.
- Términos clave.
- Ejercicios.

En este capítulo analizamos los procedimientos de prueba de hipótesis con variables cuyos valores son categorías, tales como preferencias religiosas o color de pelo. Los procedimientos a los que nos referimos se concentran en la cantidad de personas de las diferentes categorías más que en la media de alguna dimensión.

Ejemplo

Analicemos un ejemplo. Harter et al. (1997) estaban interesados en tres estilos de relaciones amorosas: un estilo autónomo concentrado en sí mismo, un estilo de relación en el que la atención está puesto en el otro, y un estilo de mutua reciprocidad. Para reunir información acerca de los estilos mencionados, realizaron una encuesta por medio de periódicos con ítems que evaluaban tanto los estilos de aquellos que respondían como la percepción de aquellos que respondían con respecto al estilo de sus parejas. Una de las predicciones del investigador establecía que los hombres que se describían a sí mismos como autónomos y concentrados en sí mismos casi seguramente describirían a sus parejas como personas que ponían su atención en el otro.

Harter y sus colegas descubrieron lo siguiente. De los 101 hombres en su estudio que se describieron como autónomos concentrados en sí mismos, el 49,5% (50 hombres) "informaron el tipo de pareja predicha, comparado con el 25,5% (26 hombres) que informaron tener parejas autónomas concentradas en sí mismas y el 24,5% (25 hombres) que informaron tener parejas con el estilo de reciprocidad...". (*p*. 156)

Supongamos que las parejas de estos hombres hubieran tenido las mismas probabilidades de tener cada uno de los tres estilos de relación. Si ese fuera el caso, entonces aproximadamente 33,66 (1/3 de los 101) de las parejas de estos hombres deberían haber pertenecido a cada uno de los tres estilos diferentes. La información que estamos manejando aparece en la segunda y tercera columna de la tabla 14-1. La segunda columna ("frecuencia observada") indica el detalle de los estilos de relación de pareja realmente observados, y la tercera columna ("frecuencia esperada") indica el detalle que se esperaría si los diferentes estilos de pareja hubieran tenido exactamente la misma probabilidad de ocurrir.

Tabla 14-1.
Frecuencias observadas y esperadas de los estilos de relación de las parejas de hombres autónomos concentrados en sí mismos.

Estilo de pareja	Frecuencia observada[a] (*O*)	Frecuencia esperada (*E*)	Diferencia (*O* – *E*)	Diferencia cuadrática $(O-E)^2$	Diferencia cuadrática ponderada según la frecuencia esperada $(O-E)^2/E$
Relación con el centro de atención puesto en el otro	50	33,67	16,33	266,67	7,92
Autónomo concentrado en sí mismo	26	33,67	–7,67	58,83	1,75
De reciprocidad	25	33,67	–8,67	75,17	2,23

[a]**Fuente:** Harter et al. (1997).

Queda claro que existe una diferencia entre lo que realmente se observó y el detalle de lo que se hubiera esperado si los estilos fueran igualmente probables. La cuestión es la siguiente: ¿Deberíamos suponer que la discrepancia observada no es más que la que esperaríamos sólo por casualidad en una muestra de este tamaño? Supongamos que las mujeres de los tres estilos tienen las mismas probabilidades de ser parejas de los hombres concentrados en sí mismos en general (la población). Aun así, en cualquier muestra en particular tomada de esa población no esperaríamos que las composiciones de los estilos de parejas fueran perfectamente iguales. Pero si la composición de la muestra está muy lejos de ser pareja, dudaríamos de que las composiciones de los estilos de pareja en la población fueran realmente iguales. En otras palabras, tenemos una situación de prueba de hipótesis muy parecida a la que hemos estado considerando hasta ahora, aunque con una diferencia importante.

En las situaciones descriptas en capítulos anteriores, los valores observados siempre han sido valores numéricos referidos a alguna dimensión, como por ejemplo, una puntuación en una prueba estándar de evaluación de nivel, de la duración de una relación, la calificación de la efectividad de un empleado por parte del empleador en una escala de 9 puntos, la cantidad de errores en la identificación de palabras, y así sucesivamente. Por el contrario, el estilo de relación de pareja de un hombre es un ejemplo de lo que en el capítulo 1 denominamos **variable nominal** (o **variable categórica**). Una variable nominal es aquella en la que la información es la cantidad de personas en cada categoría. (Se denominan variables nominales porque las diferentes categorías o niveles de la variable se identifican con nombres en lugar de números).

La prueba de hipótesis con variables nominales es una de las denominadas pruebas chi-cuadrado.[1] Las pruebas chi-cuadrado fueron desarrolladas originalmente por Karl Pearson (*véase* cuadro 14-1).

EL ESTADÍSTICO CHI-CUADRADO Y LA PRUEBA CHI-CUADRADO DE BONDAD DE AJUSTE

La idea básica de cualquier prueba chi-cuadrado es que se compara la forma con que el esquema de repartición observado de personas en varias categorías se ajusta a un esquema esperado (como

[1] Chi es la letra griega χ; se pronuncia cai.

Cuadro 14-1.
Karl Pearson, inventor del chi-cuadrado y centro de controversias.

Karl Pearson, hijo de un abogado de Yorkshire, nació en el año 1857. Pearson es muchas veces aclamado como el fundador de las ciencia estadística. La mayoría de sus virtudes y de sus vicios se revelan en lo que él relató a su colega Julia Bell como sus primeros recuerdos: estaba sentado en su sillita alta, con el pulgar en la boca, cuando le dijeron que dejara de hacerlo o si no su pulgar iba a desaparecer. Pearson miró sus dos pulgares y silenciosamente concluyó: "No veo que el pulgar que me llevo a la boca sea para nada más pequeño que el otro; me pregunto si me estarán mintiendo." Lo anterior refleja la confianza que Pearson tenía en sí mismo y en las pruebas obtenidas por la observación, como también su negación a la autoridad. También podemos observar su tendencia a dudar del carácter de las personas con quienes no estaba de acuerdo.

Pearson estudió matemática gracias a una beca en Cambridge. Poco después de ingresar, pidió que se lo excusara de las clases obligatorias de teología y del servicio religioso. Sin embargo, en cuanto accedieron a su pedido, Pearson asistió al servicio religioso. El decano lo convocó para que le diera una explicación, y Pearson declaró que no había pedido que se lo excusara del servicio religioso, "¡sino del servicio religioso obligatorio!"

Después de graduarse, Pearson viajó y estudió en Alemania, donde practicó la doctrina socialista y, como él mismo lo describía, se convirtió en un "libre pensador". Al regresar a Inglaterra, elaboró un trabajo escrito bajo un seudónimo, en el que atacaba a la cristiandad, y en 1885 fundó un club de hombres y mujeres para promover la discusión de las relaciones entre los sexos. El club desapareció, pero gracias a él conoció a su esposa, María Sharp.

Pearson finalmente se volcó a la estadística debido a su interés por probar la teoría de la evolución, y además estaba especialmente influenciado por la obra de Sir Francis Galton (*véase* cuadro 3-1). Pearson, que era mejor matemático, vio en las ideas de correlación de Galton, una forma de convertir la psicología, la antropología y la sociología en campos tan científicos como lo eran la física y la química. Esperaba evitar la cuestión de la causalidad a través de la utilización de esta categoría más amplia de correlación, asociación, o contingencia (con un rango de 0, independencia, a 1 "unidad de causalidad". "Ningún fenómeno es causal" –expresó. "Todos los fenómenos son contingentes, y el problema que enfrentamos es el de medir el grado de contingencia."

Durante toda su vida, Pearson fue muy controvertido y tuvo una fuerte voluntad, especialmente cuando se trataba de "seudo ciencia" y de la mascarada de la teología, la metafísica o las apelaciones a la autoridad bajo el pretexto de la ciencia. Incluso pensaba que la física debía dejar de utilizar palabras como **átomo, fuerza** y **materia** porque no eran fenómenos observables.

La mayor parte de su investigación, entre 1893 y 1901, se concentró en las leyes de la herencia y la evolución, pero necesitaba mejores métodos estadísticos para realizar su trabajo. Entonces se volcó a otros temas, realizando finalmente su más famosa contribución, la prueba chi-cuadrado.

Pearson también inventó el método de cálculo de la correlación utilizado en la actualidad (*véase* capítulo 3), y acuñó los términos **histograma, asimetría y correlación**

espuria. Cuando sintió que las revistas especializadas en biología no apreciaban adecuadamente su trabajo, fundó la famosa revista especializada en estadística, llamada *Biométrika*. Durante su vida, Pearson llevó la estadística de la situación de materia ampliamente ignorada a una posición primordial para el método científico, especialmente en las ciencias naturales.

Lamentablemente, Pearson era fanático de la eugenesia, el "perfeccionamiento" de la raza humana a través de la reproducción selectiva y, más tarde, su obra fue utilizada por los nazis como justificación de su trato a los judíos y otras minorías étnicas. Pero a medida que Pearson envejecía, sus opiniones enfrentaron fuerte resistencia y mucho descrédito por parte de otros estadísticos más jóvenes, lo que sólo sirvió para poner a Pearson en contra de cada vez una mayor cantidad de colegas.

De hecho, a lo largo de su vida, Pearson fue un hombre que provocó amistades devotas o, por el contrario, profunda aversión. William S. Gosset (*véase* cuadro 9-1), el inventor de la prueba *t*, fue uno de sus amigos. Sir Ronald Fisher, inventor del análisis de varianza y hombre relacionado con actitudes aún más extremas (como las descriptas en el cuadro 11-1), fue uno de los peores enemigos de Pearson (y el amable, pacífico Gosset, amigo de ambos, estaba siempre intentando suavizar los problemas entre ellos). En 1933, Pearson finalmente se retiró, y fue Fisher, nada menos, quien tomó su lugar en la cátedra de Eugenesia de Galton en la Universidad de Londres. En 1936, los dos comenzaron su más punzante discusión; Pearson murió ese mismo año.

Referencias: Peters (1987); Stiglen (1986); Tankara (1984).

por ejemplo, un esquema de repartición uniforme). Con respecto al ejemplo acerca del estilo de relación, estamos comparando el esquema observado de 50, 26 y 25 con el esquema de repartición esperado de aproximadamente 34 (33,67) para cada estilo. Un esquema de repartición de la cantidad de personas esperadas en cada categoría es, en realidad, una distribución de frecuencias como las que aprendimos en el capítulo 1. Por lo tanto, una prueba chi-cuadrado se describe más formalmente como la comparación de una distribución de frecuencias observadas con una distribución de frecuencias esperadas. En general, la prueba de hipótesis implica, primero, calcular las discrepancias entre las **frecuencias observadas** y las **frecuencias esperadas** y, después, observar si esas discrepancias son mayores de lo que se esperaría por casualidad.

Empecemos analizando de qué modo encontramos esa discrepancia entre las frecuencias observadas y esperadas. La discrepancia entre lo observado y lo esperado en cualquier categoría es simplemente la frecuencia observada menos la frecuencia esperada. Por ejemplo, veamos nuevamente el estudio de Harter et al. Con respecto a los hombres con parejas concentradas en el otro, la frecuencia observada de 50 es 16,33 puntos mayor de la frecuencia esperada de 33,67 (no debemos olvidar que la frecuencia esperada es 1/3 de 101). En la segunda categoría, la diferencia es –7,67, y en la tercera –8,67. Las diferencias mencionadas aparecen en la cuarta columna ("Diferencia") de la tabla 14-1.

Las diferencias no se utilizan directamente ya que algunas son positivas y otras negativas y, por lo tanto, se cancelarían entre sí. Este problema se resuelve elevando cada diferencia al cuadrado. (Se trata de la misma estrategia que vimos en el capítulo 2 cuando trabajamos con las diferencias de valores observados al calcular la varianza). En el ejemplo acerca del estilo de relación, la diferencia cuadrática correspondiente a parejas concentradas en el otro es de 13,33 al cuadrado, o 266,67; en el caso de las parejas concentradas en sí mismas, es de 58,83; y en el caso de las parejas con estilo de reciprocidad, 75,17. Estas diferencias cuadráticas aparecen en la quinta columna de la tabla 14-1.

En el ejemplo de Harter et al., las frecuencias esperadas son las mismas en todas las categorías. Pero en otras investigaciones, las frecuencias esperadas para las diferentes categorías pueden no ser iguales. La diferencia efectiva entre lo observado y lo esperado tiene diferente importancia según el tamaño de la frecuencia esperada. Por ejemplo, una diferencia de 8 personas, entre lo observado y lo esperado, es una discrepancia mucho mayor si la frecuencia esperada es 10 que si lo esperado es 1.000. Si la frecuencia esperada es 10, una diferencia de 8 significaría que la frecuencia observada fue de a 18 ó de 2, frecuencias tajantemente diferentes de 10. Pero si la frecuencia esperada es 1.000, una diferencia de 8 es sólo una leve desigualdad. Significaría que la frecuencia observada fue de 1.008 ó de 992, frecuencias que son sólo levemente diferentes de 1.000.

¿Cómo obtenemos un número adecuado de discrepancia (la diferencia cuadrática) entre lo observado y lo esperado con respecto a una categoría en particular? Lo que necesitamos hacer es adaptar o ponderar la desigualdad de modo tal de tener en cuenta la frecuencia esperada para esa categoría. Lo anterior se logra simplemente dividiendo la diferencia cuadrática de una categoría por la frecuencia esperada para esa categoría. Entonces, si la frecuencia esperada para determinada categoría es 10, dividimos la diferencia cuadrática por 10. Si la frecuencia esperada para la categoría es 1.000, dividimos la diferencia cuadrática por 1.000. De ese modo, ponderamos cada diferencia cuadrática según la frecuencia esperada. Esta ponderación ubica la diferencia cuadrática en una escala comparativa más adecuada.

Volvamos al ejemplo que analizábamos. En el caso de los hombres con parejas concentradas en el otro, ponderaríamos la desigualdad dividiendo la diferencia cuadrática de 266,67 por 33,67, y el resultado sería 7,92. En el caso de aquellos con parejas concentradas en sí mismas, 58,83 dividido 33,67 da 1,75; y en el caso de los hombres con parejas con estilo de reciprocidad, 75,17 dividido 33,67 da 2,23. Las desigualdades ajustadas (diferencias cuadráticas divididas por las frecuencias esperadas) aparecen en la última columna de la tabla 14-1.

Lo que resta es obtener un número general de discrepancia entre las frecuencias observadas y esperadas. Este último paso se realiza sumando los resultados de todas las categorías. Es decir, tomamos el resultado de la diferencia cuadrática dividida por la frecuencia esperada de la primera categoría, sumamos el resultado de la diferencia cuadrática dividida por la frecuencia esperada de la segunda categoría, y así sucesivamente. En el ejemplo de Harter et al. sería igual a 7,92 más 1,75 más 2,23, y daría un total de 11,90.

El número final (la suma de las diferencias cuadráticas ponderadas) es un indicador general de la discrepancia entre las frecuencias esperadas y observadas. Esa cantidad se denomina **estadístico chi-cuadrado.** Se expresa bajo la fórmula,

$$\chi^2 = \Sigma \frac{(O-E)^2}{E} \tag{14-1}$$

En la fórmula anterior, χ^2 es el estadístico chi-cuadrado. Σ es el signo de suma, que indica que debemos sumar todas las categorías distintas. O es la frecuencia observada de una categoría (la cantidad de personas realmente encontradas en esa categoría a través del estudio). E es la frecuencia esperada de una categoría (en el ejemplo que analizamos, se basa en lo que esperaríamos si hubiera la misma cantidad en todas las categorías).

Aplicando la fórmula al ejemplo de Harter et al.,

$$\chi^2 = \Sigma \frac{(O-E)^2}{E} = \frac{(50-33{,}67)^2}{33{,}67} + \frac{(26-33{,}67)^2}{33{,}67} + \frac{(25-33{,}67)^2}{33{,}67} = 11{,}90$$

Resumen de los pasos a seguir para el cálculo del estadístico chi-cuadrado

1. Encontrar las frecuencias observadas, reales, de cada categoría.
2. Determinar las frecuencias esperadas para cada categoría.
3. Calcular las frecuencias observadas menos las esperadas para cada categoría.
4. Elevar al cuadrado las diferencias de cada categoría.
5. Dividir cada diferencia cuadrática por la frecuencia esperada para cada categoría.
6. Sumar los resultados del paso 5 de todas las categorías.

La distribución chi-cuadrado

El siguiente paso es averiguar si el estadístico chi-cuadrado que hemos calculado representa una discrepancia mayor a la que podría ocurrir por casualidad. Para responder esta pregunta necesitamos saber cuáles son las probabilidades de que el chi-cuadrado tome valores de distintos intervalos por casualidad. Es decir, necesitamos la distribución del estadístico chi-cuadrado que ocurriría por casualidad. Sucede que siempre que el estudio tenga una cantidad razonable de personas, la distribución del estadístico chi-cuadrado es bastante próxima a una distribución matemática conocida que se denomina, por supuesto, **distribución chi-cuadrado.**

La manera exacta de la distribución chi-cuadrado depende de los grados de libertad. En una prueba chi-cuadrado, los grados de libertad son la cantidad de categorías que son libres de variar en cuanto a sus frecuencias, dándose como conocido el total de participantes. En el ejemplo acerca del estilo de relación hay tres categorías. Si conocemos la cantidad total de personas y también sabemos la cantidad que corresponde a dos de las categorías, automáticamente podemos calcular la cantidad de participantes en la tercera categoría. En un estudio como el del ejemplo que estamos analizando, si hay tres categorías, hay dos grados de libertad.

La figura 14-1 indica las distribuciones chi-cuadrado para varios grados de libertad. Según se observa en la figura, las distribuciones son todas asimétricas hacia la derecha. Esto se debe a que el chi-cuadrado no puede ser menor a 0, pero puede tener valores muy altos. (El chi-cuadrado debe ser positivo porque se calcula sumando un grupo de fracciones en las que el numerador y el denominador deben ser todos positivos. El numerador necesariamente es positivo porque está elevado al cuadrado, y el denominador necesariamente es positivo porque la cantidad de personas esperadas en determinada categoría no puede ser negativa, ¡no se puede esperar que haya menos que ninguna persona!).

La tabla chi-cuadrado

Lo más importante acerca de la distribución chi-cuadrado para una prueba de hipótesis es el punto de corte que indica que un chi-cuadrado es lo suficientemente grande como para rechazar la hipótesis nula. Por ejemplo, supongamos que queremos utilizar el nivel de significación de 0,05. En ese caso, necesitamos saber qué punto de la distribución chi-cuadrado tiene el 5% de los chi-cuadrados por encima de sí mismo. Una **tabla chi-cuadrado** proporciona los puntos de corte para los distintos niveles de significación y para varios grados de libertad. La tabla 14-2 muestra una parte de una tabla chi-cuadrado como la que aparece en el apéndice B (tabla B-4). Analicemos el ejemplo referido al estilo de relación, en el que había dos grados de libertad. La tabla muestra que el punto de corte chi-cuadrado para un nivel de 0,05, utilizando una distribución chi-cuadrado con 2 grados de libertad, es 5,992.

La prueba chi-cuadrado de bondad de ajuste

Ahora contamos con toda la información necesaria para probar la hipótesis en el ejemplo de Harter et al. Cabe recordar que el chi-cuadrado que calculamos para el ejemplo era de 11,90. Además, acabamos de encontrar el punto de corte correspondiente al ejemplo (utilizando el nivel 0,05 de significación), que es de 5,992. Comparando los dos números mencionados anteriormente, el chi-cuadrado del estudio es claramente superior al punto de corte. Por lo tanto, los investigadores que realizaron el estudio rechazaron la hipótesis nula, es decir, la rechazaron por considerar demasiado improbable que la discrepancia que observaron pudiera haber ocurrido si, de hecho, la población de hombres concentrados en sí mismos tuviera una cantidad igual de parejas de cada estilo de relación. Parecía más razonable sostener que los estilos de relación de las parejas de ese tipo de hombres eran realmente diferentes.

Acabamos de realizar un procedimiento de prueba de hipótesis completo del ejemplo de Harter et al. El ejemplo incluía diferentes cantidades de personas en tres niveles de una determinada variable nominal (el estilo de relación de las parejas de hombres concentrados en sí mismos). Este tipo de pruebas chi-cuadrado, que incluye niveles de una sola variable nominal, se denomina **prueba chi-cuadrado de bondad de ajuste**. (Más adelante, en el capítulo, analizaremos situaciones que incluyen más de una variable nominal a la vez).

Pasos de la prueba de hipótesis: ejemplo

Reveamos el proceso de realización de una prueba chi-cuadrado de bondad de ajuste. Utilizaremos el mismo ejemplo, pero esta vez seguiremos sistemáticamente los cinco pasos estándar. A lo largo del proceso, también analizaremos algunos detalles.

1. **Replantear el problema en función de hipótesis de investigación e hipótesis nula de las poblaciones**. Las dos poblaciones son:

Población 1: hombres concentrados en sí mismos como los que intervienen en el estudio.
Población 2: hombres concentrados en sí mismos cuyas parejas pertenecen en igual cantidad a los tres estilos de relaciones.

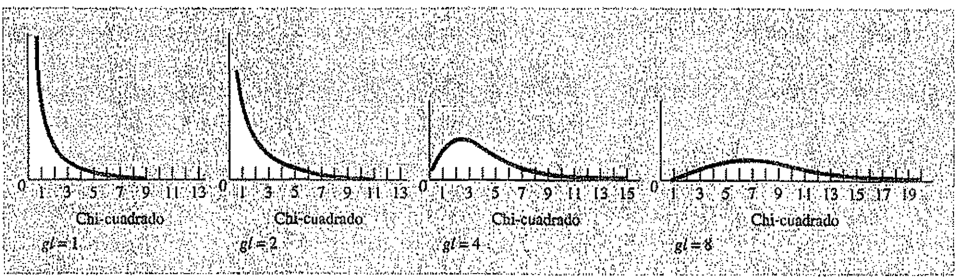


Figura 14-1.
Ejemplos de distribuciones chi-cuadrado para diferentes grados de libertad.

Tabla 14-2.
Parte de una tabla chi-cuadrado.

	Nivel de significación		
gl	0,10	0,05	0,01
1	2,706	3,841	6,635
2	4,605	**5,992**	9,211
3	6,252	7,815	11,345
4	7,780	9,488	13,277
5	9,237	11,071	15,087

La hipótesis de investigación establece que la distribución de las personas en las categorías de las dos poblaciones es diferente; la hipótesis nula establece que es igual.

2. **Determinar las características de la distribución comparativa.** La distribución comparativa en este caso es una distribución chi-cuadrado con dos grados de libertad. (Una vez que conocemos el total, sólo las cantidades en dos categorías pueden variar libremente).

Es importante no confundirnos con la terminología. La distribución comparativa es la distribución con la que comparamos el número que resume todo el patrón del resultado. Con una prueba *t*, este número es el punto *t*, y utilizamos una distribución *t*. Con un análisis de varianza, es la razón *F*, y utilizamos una distribución *F*. Del mismo modo, con una prueba chi-cuadrado, la distribución es una distribución chi-cuadrado.

Decimos que puede surgir cierta confusión, ya que al prepararnos para utilizar la distribución chi-cuadrado comparamos una distribución de frecuencias observadas con una distribución de frecuencias esperadas. Pero la distribución de frecuencias esperadas no es una distribución comparativa en el sentido en el que utilizamos ese término en el paso 2 de la prueba de hipótesis.

3. **Determinar el punto de corte en la distribución comparativa, a partir del cual debería rechazarse la hipótesis nula.** Buscamos el punto de corte en la tabla chi-cuadrado según el nivel de significación y los grados de libertad del estudio. En este caso, utilizamos el nivel 0,05 de significación y determinamos, en el paso 2, que había 2 grados de libertad. Basándonos en la tabla, el chi-cuadrado de corte es igual a 5,992.

4. **Determinar el valor muestral en la distribución comparativa.** El valor muestral es el chi-cuadrado calculado a partir de la muestra. En otras palabras, este es el paso en el que se realizan todos los cálculos; es decir, para cada categoría necesitamos calcular las frecuencias esperadas, las diferencias entre las frecuencias esperadas y observadas elevadas al cuadrado, y dividir ese resultado por la frecuencia esperada. Sumando los resultados de todos estos cálculos para cada categoría obtenemos el chi-cuadrado del estudio. En el ejemplo que estamos utilizando el resultado es 11,90.

5. Comparar los valores obtenidos en los pasos 3 y 4 para decidir si se rechaza o no la hipótesis nula. Dado que el punto de corte para rechazar la hipótesis nula es 5,992 y el chi-cuadrado de nuestra muestra es 11,90, podemos rechazar la hipótesis nula. Se sostiene la hipótesis de investigación que establece que las dos poblaciones son diferentes. Es decir, los investigadores concluyen que las parejas de hombres concentrados en sí mismos no tienen las mismas probabilidades de presentar los tres estilos de relación.

Otro ejemplo

Analicemos otro ejemplo. Un equipo de investigación ficticio formado por psicólogos clínicos desea probar una teoría que establece que la salud mental se ve afectada por el nivel de cierto mineral incluido en la dieta alimenticia. Al mineral lo llamaremos Q. El equipo de investigación ha localizado una región de los Estados Unidos cuyo suelo presenta una alta concentración del mineral Q y, debido a ello, ese mineral se encuentra en el agua que las personas consumen y en los alimentos que se siembran en el lugar. Los investigadores realizan una encuesta a personas mayores que han vivido toda su vida en esa área, concentrándose en los trastornos de la salud mental. De las 1.000 personas entrevistadas, 134 habían experimentado en algún momento de su vida un trastorno relacionado con la angustia, 160 habían sufrido alcoholismo o drogadicción, 97 trastornos de estados anímicos (tales como depresión crónica) y 12 habían sufrido esquizofrenia; 597 nunca habían experimentado ninguno de los problemas anteriores. (En este ejemplo, ignoraremos lo que ocurre cuando una persona ha sufrido más de uno de los trastornos).

Los psicólogos compararon los resultados con lo que se esperaría sobre la base de una gran encuesta realizada al público en general de los Estados Unidos. En esa encuesta, el 14,6% de los adultos en algún momento de sus vidas sufre de trastornos relacionados con la angustia, el 16,4% padece alcoholismo y drogadicción, el 8,3% sufre trastornos del estado anímico y el 1,5% padece esquizofrenia, mientras que el 59,2% no experimenta ninguno de esos trastornos (Regier et al., 1984). Si la muestra de 1.000 no es diferente de la población general de Estados Unidos, el 14,6% de ellos (146) deberían haber sufrido trastornos relacionados con la angustia, el 16,4% (164) deberían haber padecido alcoholismo y drogadicción, y así sucesivamente. La cuestión planteada por los psicólogos clínicos es la siguiente: sobre la base de la muestra que hemos estudiado, ¿podemos concluir que los porcentajes de los diferentes problemas mentales sufridos por las personas de esta región son diferentes a los de la población de los EE.UU. en general?

La tabla 14-3 indica las frecuencias observadas y esperadas y los cálculos de la prueba chi-cuadrado.

1. **Replantear el problema en función de hipótesis de investigación e hipótesis nula de las poblaciones.** Las dos poblaciones son:

Población 1: las personas de la región de EE.UU. con alto nivel del mineral Q.
Población 2: la población de EE.UU.

La hipótesis de investigación establece que la distribución de cantidades de personas, entre las cinco categorías de salud mental, es diferente en las dos poblaciones; la hipótesis nula establece que es igual.

2. **Determinar las características de la distribución comparativa.** La distribución comparativa es una distribución chi-cuadrado con 4 grados de libertad (es decir, 5 categorías − 1 = 4).

3. **Determinar el punto de corte en la distribución comparativa, a partir del cual debería rechazarse la hipótesis nula.** Utilizaremos el nivel de significación estándar del 5%, y hay, según vimos, 4 grados de libertad. Por lo tanto, la tabla 14-2 (o la tabla B-4 del apéndice B) indica que los psicólogos clínicos necesitan un chi-cuadrado de al menos 9,488 para rechazar la hipótesis nula. La figura 14-2 representa la situación gráficamente.

4. **Determinar el valor muestral en la distribución comparativa.** La tabla 14-3 indica los cálculos del chi-cuadrado. Se sigue el procedimiento usual: encontrar la diferencia entre las frecuencias observadas y esperadas de cada categoría, elevarlas al cuadrado, dividir cada una por la cantidad esperada para cada categoría, y luego sumar los resultados de las distintas categorías. El resultado es un chi-cuadrado de 4,09.

Tabla 14-3.
Frecuencias observadas y esperadas y prueba chi-cuadrado de la bondad de ajuste de distintos tipos de trastornos de la salud mental en una región de EEUU con alto nivel del mineral Q, comparada con la población de EE.UU. en general. (Datos ficticios).

Condición	Observada	Esperada
Angustia	134	146 (14,6% × 1.000)
Alcoholismo y drogadicción	160	164 (16,4% × 1.000)
Trastornos del estado anímico	97	83 (8,3% × 1.000)
Esquizofrenia	12	15 (1,5% × 1.000)
Ninguna de las anteriores	597	592 (59,2% × 1.000)

Grados de libertad = 5 categorías − 1 = 4

Punto de corte chi-cuadrado necesario, $gl = 4$, nivel 0,05: 9,488

$$\chi^2 = \Sigma \frac{(O-E)^2}{E} = \frac{(134-146)^2}{146} + \frac{(160-164)^2}{164} + \frac{(97-83)^2}{83} + \frac{(12-15)^2}{15} + \frac{(597-592)^2}{592}$$

$$= \frac{-12^2}{146} + \frac{-4^2}{164} + \frac{14^2}{83} + \frac{-3^2}{15} + \frac{5^2}{592} = \frac{144}{146} + \frac{16}{164} + \frac{196}{83} + \frac{9}{15} + \frac{25}{592}$$

$$= 0{,}99 + 0{,}10 + 2{,}36 + 0{,}60 + 0{,}04 = 4{,}09$$

Conclusión: no se rechaza la hipótesis nula.

5. **Comparar los valores de los pasos 3 y 4 para decidir si se rechaza o no la hipótesis nula.** El chi-cuadrado de 4,09 es mucho menos extremo que el punto de corte de 9,488 (*véase* figura 14-2). Los investigadores no pueden rechazar la hipótesis nula; el estudio no es concluyente. (No habiendo podido rechazar la hipótesis nula con una muestra tan grande, es razonable suponer que si existe alguna diferencia entre las poblaciones, esa diferencia es bastante pequeña).

Un tercer ejemplo

Supongamos que una profesora de una gran universidad está dando un curso de introducción a la estadística a 200 alumnos. La clase ya ha terminado de rendir su parcial. Anteriormente, la profesora siempre ha calificado con una curva aproximada a la distribución normal, es decir, el 2,5% superior de los alumnos obtuvo A, el siguiente 14% recibió B, el siguiente 67% recibió C, el siguiente 14% recibió D y el 2,5% más bajo recibió F.

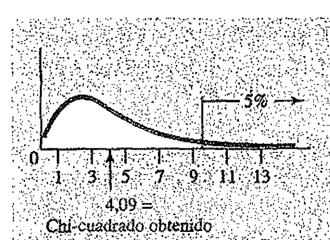


Figura 14-2.
Distribución chi-cuadrado ($gl = 4$) correspondiente al ejemplo del mineral Q, que muestra el punto de corte para el rechazo de la hipótesis nula al nivel 0,05.

Este año, sin embargo, la profesora ha decidido asignar las calificaciones según el porcentaje del examen realizado correctamente; un 90% ó más es una A, entre un 80% y 89% una B, y así sucesivamente. La pregunta que la profesora se plantea entonces es la siguiente: sobre la base de la muestra de este semestre formada por 200 calificaciones de parciales a través del nuevo sistema, ¿existe alguna razón para creer que el nuevo sistema produce una distribución diferente de calificaciones?

La tabla 14-4 indica las frecuencias observadas y esperadas y los cálculos de la prueba chi-cuadrado.

1. **Replantear el problema en función de hipótesis de investigación e hipótesis nula de las poblaciones.** Las dos poblaciones son:

Población 1: alumnos calificados según el nuevo sistema (que tiene en cuenta sus calificaciones sin importar el nivel de los otros alumnos de la clase).
Población 2: alumnos calificados con una curva de distribución normal.

La hipótesis de investigación establece que las poblaciones son diferentes; la hipótesis nula establece que las poblaciones son iguales.

2. **Determinar las características de la distribución comparativa.** La distribución comparativa es una distribución chi-cuadrado con 4 grados de libertad (5 categorías −1 = 4).

3. **Determinar el punto de corte en la distribución comparativa, a partir del cual se debería rechazar la hipótesis nula.** La profesora es conservadora en cuanto a sus decisiones estadísticas y, por lo tanto, elige el nivel 0,01. Utilizando la tabla 14-2 (o la tabla B-4) para 4 grados de libertad, el profesor necesita un chi-cuadrado de al menos 13,277 para rechazar la hipótesis nula.

Tabla 14-4.
Frecuencias observadas y esperadas y prueba chi-cuadrado de bondad de ajuste del ejemplo de calificaciones parciales. (Datos ficticios).

Calificación	Observada	Esperada
A	10	5 (2,5% × 200)
B	34	28 (14,0% × 200)
C	140	134 (67,0% × 200)
D	10	28 (14,0% × 200)
F	6	5 (2,5% × 200)

Grados de libertad = 5 categorías − 1 = 4

Punto de corte chi-cuadrado necesario, $gl = 4$, nivel 0,01: 13,277

$$\chi^2 = \Sigma \frac{(O-E)^2}{E} = \frac{(10-5)^2}{5} + \frac{(34-28)^2}{28} + \frac{(140-134)^2}{134} + \frac{(10-28)^2}{28} + \frac{(6-5)^2}{5}$$

$$= \frac{5^2}{5} + \frac{6^2}{28} + \frac{6^2}{134} + \frac{-18^2}{28} + \frac{1^2}{5} = \frac{25}{5} + \frac{36}{28} + \frac{36}{134} + \frac{324}{28} + \frac{1}{5}$$

$$= 5 + 1{,}29 + 0{,}27 + 11{,}57 + 0{,}20 = 18{,}33$$

Conclusión: Se rechaza la hipótesis nula.

4. **Determinar el valor muestral en la distribución comparativa.** Para calcular el chi-cuadrado, primero calculamos las frecuencias esperadas multiplicando los porcentajes esperados por la cantidad en la muestra. Para el primer grupo (calificación A), la profesora esperaba un 2,5% según el sistema de curva normal que había utilizado previamente: 2,5% x 200 = 5. Por lo tanto, para las calificaciones A, ella esperaba una frecuencia de 5. Según el antiguo sistema, el 14% habría obtenido una B, lo que da una frecuencia esperada de 28 alumnos de su clase de 200. La tabla 14-4 indica el resto de las frecuencias esperadas más los cálculos del chi-cuadrado. Como se observa en la tabla, el resultado es un chi-cuadrado de 18,33.

5. **Comparar los valores obtenidos en los pasos 3 y 4 para determinar si se rechaza o no la hipótesis nula.** El chi-cuadrado necesario para rechazar la hipótesis nula era 13,277. El chi-cuadrado de la muestra es 18,33. Por lo tanto, la profesora puede rechazar la hipótesis nula y concluir que las poblaciones son diferentes (*véase* figura 14-3). El nuevo método de calificación no produjo una distribución normal de las calificaciones de la clase. Si bien no se predijo la dirección de la diferencia, un análisis de los valores de las categorías muestra que, en este ejemplo, utilizar el método de calificación por puntos dio como resultado que más alumnos obtuvieran A, B ó C, y menos alumnos obtuvieran D y F.

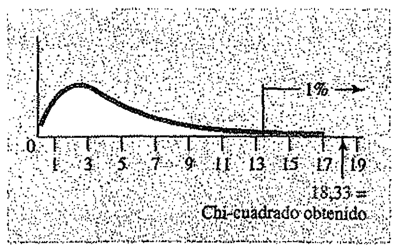
PRUEBA CHI-CUADRADO DE INDEPENDENCIA

Hasta ahora hemos observado la distribución de una variable nominal con varias categorías, tales como estilos de relación de las parejas de los hombres. De hecho, es bastante raro que este tipo de situación ocurra en la investigación. Comenzamos con un ejemplo de ese tipo principalmente porque es un buen escalón para llegar a la situación más común de investigación real, la situación que trataremos a continuación.

La situación más común en la que se utiliza el chi-cuadrado es aquella en la que existen dos variables nominales, cada una con varias categorías. Por ejemplo, Harter et al. podrían haber estado interesados en saber si la clasificación de las parejas de hombres concentrados en sí mismos era diferente de la clasificación de las parejas de hombres concentrados en el otro u hombres con estilo de reciprocidad. Si ese hubiera sido su propósito, habríamos tenido dos variables nominales. El estilo de relación de las parejas habría sido la primera variable nominal, y el estilo de relación propio de los hombres habría sido la segunda variable nominal. La prueba de hipótesis, en este tipo de situaciones, se denomina **prueba chi-cuadrado de independencia.** Pronto veremos el porqué de este nombre.

Analicemos el siguiente estudio ficticio. Los investigadores de una gran universidad realizan una encuesta a 200 miembros del personal que viajan diariamente a su trabajo. A los miembros

Figura 14-3.
Distribución chi-cuadrado (gl = 4) del ejemplo acerca del sistema de calificaciones, que muestra el punto de corte para el rechazo de la hipótesis nula al nivel 0,01.



del personal se les pregunta acerca del tipo de transporte que utilizan, y si prefieren acostarse temprano y levantarse temprano ("personas diurnas") o acostarse tarde y levantarse tarde ("personas nocturnas"). Los resultados se reflejan en la tabla 14-5. Observemos las dos variables nominales: a) tipo de transporte, con tres niveles y b) tendencia de descanso, con dos niveles.

Tablas de contingencia

La tabla 14-5 es un ejemplo de una **tabla de contingencia**, aquella en la que se establecen las distribuciones de dos variables nominales de modo que refleje las frecuencias de sus combinaciones y también los totales. Una tabla de contingencia es similar a las tablas utilizadas en los diseños factoriales de investigación que se analizan con un análisis de varianza de dos criterios (*véase* capítulo 13). Sin embargo, en una tabla de contingencia, los números son frecuencias y no medias. El número en cada categoría o combinación de categorías es una cantidad de individuos, no un promedio de registros de determinada clase. Por lo tanto, en la tabla 14-5, el 60 en la combinación autobús-diurna expresa cuántas personas diurnas toman autobús. No es un promedio ni nada que se le parezca.

La tabla 14-5 es un ejemplo de tabla de contingencia 3 x 2 porque tiene tres niveles de una variable cruzados con dos niveles de otra variable (no importa qué dimensión se nombre primero). También es posible crear tablas de contingencia mayores, como por ejemplo de 4 x 7 ó 6 x 18. Las tablas más pequeñas, las tablas de contingencia 2 x 2, son las más comunes.

Independencia

El objetivo en el ejemplo que estamos analizando es saber si existe alguna relación entre el tipo de transporte que utilizan las personas y el hecho de que sean personas diurnas o nocturnas. Si no existe relación, la proporción de personas diurnas y nocturnas será la misma entre los que viajan en autobús, los que comparten los autos y los que van con sus propios autos. O para decirlo de otro modo, si no existe relación, la proporción de personas que viajan en autobús –comparten los autos y conducen sus propios autos–, es la misma en el caso que se trate de personas diurnas o nocturnas. No importa cómo se describa. La situación de ausencia de relación entre las variables en una tabla de contingencia se denomina independencia.[2]

Tabla 14-5.
Tabla de contingencia de frecuencias observadas de personas diurnas y nocturnas que utilizan diferentes medios de transporte. (Datos ficticios).

		Medio de transporte			**Total**
		Autobús	*Automóvil compartido*	*Automóvil propio*	
Tendencia de descanso	*Diurna*	60	30	30	120 (60%)
	Nocturna	20	20	40	80 (40%)
	Total	80	50	70	200 (100%)

[2] El término independencia se utiliza usualmente para referirse a la ausencia de relación entre dos variables nominales. Sin embargo, si el alumno ya ha estudiado el capítulo 3, puede resultarle útil pensar en la independencia como algo similar a la situación de falta de correlación o coeficiente de correlación 0 ($r = 0$).

Muestra y población

Según los resultados de la encuesta observados en el estudio, las proporciones de personas nocturnas y diurnas de la muestra varían de acuerdo con los diferentes medios de transporte. Por ejemplo, los que viajan en autobús se dividen en 60-20, es decir, tres cuartas partes de los que viajan en autobús son personas diurnas. Entre las personas que viajan en su propio auto, la división es 30-40, es decir, una leve mayoría son personas nocturnas. Aun así, debemos tener en cuenta que la muestra es de sólo 200 personas, y es posible que en la población mayor, el tipo de transporte que utiliza una persona sea independiente del hecho de que esa persona sea diurna o nocturna. La gran pregunta es si la falta de independencia en la muestra es lo suficientemente grande *como para rechazar la hipótesis nula de independencia* en la población.

Utilización del chi-cuadrado en una prueba de independencia

Para probar si la falta de independencia en una muestra es lo suficientemente grande como para rechazar la hipótesis nula de independencia en la población, se requieren dos elementos. En primer lugar, necesitamos un número que refleje la desigualdad entre el patrón de la muestra y lo que esperaríamos si el patrón de la muestra reflejara perfectamente una población en la que hubiera independencia. Ese número es el estadístico chi-cuadrado. En segundo lugar, necesitamos conocer la distribución de ese estadístico si la hipótesis nula fuera verdadera, es decir, la *distribución chi*-cuadrado.

Tal como hicimos en el ejemplo referido al estilo de relación, debemos calcular un chi-cuadrado y compararlo con un punto de corte chi-cuadrado tomado de una tabla. Lo nuevo en este caso son los detalles en cuanto a la forma de calcular el chi-cuadrado y en cuanto a la forma de calcular los grados de libertad para buscar el punto de corte en la tabla chi-cuadrado.

Determinación de las frecuencias esperadas

Tal como hicimos anteriormente, para calcular el chi-cuadrado comparamos las frecuencias observadas con las esperadas. La novedad en este caso es que ahora tenemos que calcular diferencias entre lo observado y lo esperado para cada combinación de categorías, es decir, para cada casilla de la tabla de contingencia (cuando había sólo una variable nominal, calculábamos estas diferencias sólo para cada categoría de la única variable nominal). La novedad más importante del procedimiento está relacionada con el cálculo de lo que deberían ser las frecuencias esperadas.

La tabla 14-6 es la tabla de contingencia correspondiente a la encuesta del ejemplo. Esta vez hemos incluido la frecuencia esperada (entre paréntesis) al lado de cada frecuencia observada. Es recomendable ir analizando la lógica que indican los dos párrafos siguientes mientras se observan los números mencionados.

Para calcular frecuencias esperadas suponemos que las dos variables son independientes, es decir, en este ejemplo, suponemos que el transporte y la tendencia de descanso son independientes (suponemos esto al calcular las frecuencias esperadas porque es con la independencia con lo que queremos comparar nuestras frecuencias observadas). Si son independientes, entonces las proporciones entre las casillas superiores e inferiores de cada columna de transporte deberían ser iguales. Por ejemplo, la proporción de personas diurnas que viajan en autobús debería ser igual a la proporción de personas diurnas entre aquellos que comparten los autos, e igual a la proporción de personas diurnas que utilizan su propio auto. De hecho, todas estas proporciones deberían ser iguales a la proporción de personas diurnas en la encuesta en general. Para decirlo de otro modo,

Tabla 14-6.
Tabla de contingencia de frecuencias observadas (y esperadas) de personas diurnas y nocturnas que utilizan diferentes tipos de transporte. (Datos ficticios).

		Medio de transporte			Total
		Autobús	*Automóvil compartido*	*Automóvil propio*	
Tendencia de descanso	*Diurna*	60 (48)[a]	30 (30)	30(42)	120 (60%)
	Nocturna	20(32)	20(20)	40(28)	80 (40%)
	Total	80	50	70	200 (100%)

[a] Las frecuencias esperadas están entre paréntesis.

el patrón proporcional de personas diurnas y nocturnas en cada columna debería ser igual que al de toda la distribución. Lo anterior significaría que el medio de transporte no afecta la proporción de personas diurnas y nocturnas, y que el medio de transporte es independiente de la proporción de personas diurnas y nocturnas.

Analicemos ahora los números reales de la encuesta del ejemplo. En total existe un 60% de personas diurnas y un 40% de personas nocturnas. Por lo tanto, si el medio de transporte es independiente del hecho de ser una persona diurna o nocturna, este 60%-40% debería mantenerse en cada columna (cada tipo de transporte). En primer lugar, el 60%-40% total debería mantenerse en el grupo de personas que viajan en autobús, es decir, que en la casilla de la personas diurnas que toman el autobús esperaríamos una frecuencia del 60% sobre 80, es decir, 48 personas. La frecuencia esperada para las personas nocturnas que toman autobús es 32 (es decir, el 40% de 80 es 32). Del mismo modo, analicemos las frecuencias esperadas para la columna de aquellos que comparten el automóvil. La columna debería dividirse en 60%-40%; por lo tanto, se espera que su total de 50 personas se divida en un 60% - 40%, dando como resultado una frecuencia esperada de 30 personas diurnas que viajan en automóviles compartidos (es decir, el 60% de 50 es 30) y 20 personas nocturnas que viajan en automóviles compartidos (es decir, el 40% de 50 es 20). Las frecuencias esperadas para la columna de personas que viajan en sus propios automóviles se calculan del mismo modo, y dan 42 y 28, tal como lo muestra la tabla 14-6.

Lo anterior se expresa bajo la fórmula,

$$E = \left(\frac{R}{N}\right)(C) \qquad (14\text{-}2)$$

En la fórmula, E es la frecuencia esperada para una casilla en particular (la combinación de categorías); R es la cantidad de personas observadas en la fila de esa casilla; N es la cantidad total de personas, y C es la cantidad de personas observadas en la columna de esa casilla. (Aun si se confunden las casillas y las columnas, la frecuencia esperada resulta la misma).

Aplicando la fórmula a las personas diurnas que viajan en autobús,

$$E = \left(\frac{R}{N}\right)(C) = \left(\frac{120}{200}\right)(80) = (\,0{,}60)(80) = 48$$

Observando la tabla 14-6 en su totalidad, vemos que las frecuencias esperadas suman los mismos totales de columnas y filas que las frecuencias observadas. Por ejemplo, en la primera columna (autobús), las frecuencias esperadas de 32 y 48 suman 80, al igual que las frecuencias observadas de 60 y 20 de la misma columna. De modo similar, en la fila superior (diurna), las frecuencias esperadas de 48, 30 y 42 suman 120, el mismo total de las frecuencias observadas de 60, 30 y 30. Para controlar los cálculos aritméticos, es siempre una buena idea asegurarse de que las frecuencias esperadas y observadas sumen los mismos totales tanto de fila como de columna.

Cálculo del chi-cuadrado

Una vez que conocemos las frecuencias observadas y esperadas, podemos calcular el chi-cuadrado del mismo modo que en la prueba chi-cuadrado de bondad de ajuste. Los pasos y las fórmulas son exactamente las mismas. La única diferencia es que ahora calculamos la diferencia cuadrática ponderada para cada **casilla** y luego las sumamos (antes hicimos lo mismo para cada categoría y no había casillas para combinaciones de categorías porque había sólo una variable nominal). Abajo indicamos cómo funciona el proceso aplicado a la encuesta del ejemplo:

$$\chi^2 = \Sigma \frac{(O-E)^2}{E}$$

$$= \frac{(60-48)^2}{48} + \frac{(30-30)^2}{30} + \frac{(30-42)^2}{42} + \frac{(20-32)^2}{32} + \frac{(20-20)^2}{20} + \frac{(40-28)^2}{28}$$

$$= 3 + 0 + 3{,}43 + 4{,}5 + 0 + 5{,}14 = 16{,}07$$

Grados de libertad

Como siempre, antes de que podamos probar la significación debemos saber cuáles son los grados de libertad. Los grados de libertad para el chi-cuadrado de una tabla de contingencia son la cantidad de columnas menos 1 por la cantidad de filas menos 1. Se expresa bajo la fórmula,

$$gl = (N_{\text{Columnas}} - 1)(N_{\text{Filas}} - 1) \qquad (14\text{-}3)$$

En la fórmula anterior, N_{Columnas} es la cantidad de columnas y N_{Filas} es la cantidad de filas. Si aplicamos esta fórmula a la encuesta del ejemplo,

$$gl = (N_{\text{Columnas}} - 1)(N_{\text{Filas}} - 1) = (3-1)(2-1) = (2)(1) = 2$$

Una tabla de contingencia con muchas casillas puede tener relativamente pocos grados de libertad, ya que en una prueba chi-cuadrado los grados de libertad son la cantidad de categorías libres de variar una vez que se conocen los totales. Con una prueba chi-cuadrado de independencia, la cantidad de categorías se convierte en la cantidad de casillas; los totales ahora incluyen totales de fila y columna al igual que el total general. Si conocemos los totales de fila y de columna, contamos con mucha información.

Analicemos el ejemplo acerca de la tendencia de descanso y los medios de transporte. Si conocemos las frecuencias de las primeras dos casillas superiores, por ejemplo, y todos los totales de fila y columna, podríamos calcular todas las otras frecuencias. La tabla 14 -7 muestra la tabla de contingencia correspondiente al ejemplo que analizamos, con sólo los totales de fila y columna (y el total general) y las frecuencias de esas dos casillas. Podemos completar el resto de la fila superior calculando que, si hay un total de 120 (el total de esa fila) y las otras dos casillas dan 90 entre las dos (60 + 30), entonces sólo quedan 30 que se ubican en la casilla de automóvil propio. Si co-

Tabla 14-7.
Tabla de contingencia que incluye las frecuencias observadas marginales y de dos casillas, con el fin de ilustrar el cálculo de los grados de libertad.

		Medio de transporte			Total
		Autobús	*Automóvil compartido*	*Automóvil propio*	
Tendencia de descanso	*Diurna*	60	30	—	120 (60%)
	Nocturna	—	—	—	80 (40%)
	Total	80	50	70	200 (100%)

nocemos las frecuencias de todas las casillas de personas diurnas y los totales de columnas para cada tipo de transporte, entonces la frecuencia de cada casilla correspondiente a las personas nocturnas es igual al total de su columna menos las personas diurnas de esa columna. Por ejemplo, hay 80 personas que viajan en autobús y 60 son personas diurnas, entonces los 20 restantes deben ser personas nocturnas. Por lo tanto, en este ejemplo, aunque hay seis celdas, hay sólo 2 grados de libertad; entonces, hay sólo dos casillas cuyas frecuencias son realmente libres de variar una vez que tenemos todos los totales de fila y columna.

Prueba de hipótesis

Con 2 grados de libertad, la tabla 14-2 (o tabla B-4) muestra que el punto de corte chi-cuadrado necesario para tener significación a un nivel de 0,01 es 9,211. El chi-cuadrado de 16,07 del ejemplo es mayor que ese punto de corte. Por lo tanto, podemos rechazar la hipótesis nula que establece que en la población las dos variables son independientes.

Pasos de la prueba de hipótesis y prueba chi-cuadrado de independencia: un ejemplo

Acabamos de realizar una prueba de hipótesis completa utilizando la prueba chi-cuadrado de independencia. Sin embargo, una vez más será útil rever el proceso, pero esta vez siguiendo sistemáticamente los cinco pasos de la prueba de hipótesis.

1. **Replantear el problema en función de hipótesis de investigación e hipótesis nula de las poblaciones.** Las dos poblaciones son:

Población 1: personas como las entrevistadas.
Población 2: personas para las cuales ser nocturna o diurna es independiente del tipo de transporte utilizado para ir a trabajar.

La hipótesis nula establece que las dos poblaciones son iguales, y que en general las proporciones que utilizan diferentes tipos de transporte son las mismas para las personas diurnas y nocturnas. La hipótesis de investigación establece que las dos poblaciones son diferentes, y que entre las personas en general, las proporciones que utilizan diferentes tipos de transporte varían según se trate de personas diurnas o nocturnas.

Para decirlo de otro modo, la hipótesis nula establece que las dos variables son independientes (no están relacionadas entre sí). La hipótesis de investigación establece que no son independientes (que están relacionadas entre sí).

2. **Determinar las características de la distribución comparativa.** La distribución comparativa es una distribución chi-cuadrado con 2 grados de libertad. Si conocemos la cantidad de participantes de dos casillas y los totales de fila y columna, todas las demás cantidades pueden determinarse. O bien, utilizando la regla para tablas de contingencia, la cantidad de casillas libres de variar es la cantidad de columnas menos 1 por la cantidad de filas menos 1.

3. **Determinar el punto de corte en la distribución comparativa, a partir del cual debería rechazarse la hipótesis nula.** Utilizamos la misma tabla que para cualquier prueba chi-cuadrado. En el ejemplo, estableciendo un nivel de 0,01 de significación con 2 grados de libertad, necesitamos un chi-cuadrado de 9,211.

4. **Determinar el valor muestral en la distribución comparativa.** En el ejemplo, encontramos un chi-cuadrado de 16,07.

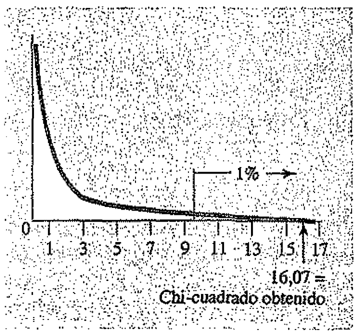
5. **Comparar los valores obtenidos en los pasos 3 y 4 para determinar si se rechaza o no la hipótesis nula.** El chi-cuadrado necesario para rechazar la hipótesis nula es de 9,211, y el chi-cuadrado del ejemplo es de 16,07 (*véase* figura 14-4). Por lo tanto, podemos rechazar la hipótesis nula. Se sostiene la hipótesis de investigación que establece que, en la población, las dos variables no son independientes. En consecuencia, las proporciones del tipo de transporte utilizado para ir a trabajar difiere según se trate de personas diurnas o nocturnas.

Un segundo ejemplo

En el año 1994, Riehl realizó un estudio para analizar la experiencia universitaria de alumnos de primer año que eran la primera generación de la familia en asistir a la universidad. Los alumnos fueron comparados con otros alumnos que no eran la primera generación de la familia que asistía a la universidad (todos los alumnos pertenecían a la Universidad de Indiana). Una de las variables que midió Riehl fue si los alumnos abandonaban o no los estudios durante el primer semestre.

La tabla 14-8 muestra los resultados y los porcentajes correspondientes a los grupos de abandono y no abandono, más las frecuencias esperadas (entre paréntesis) basadas en esos porcentajes. Debajo de la tabla de contingencia se encuentran los cálculos de la prueba chi-cuadrado de independencia.

Figura 14-4.
Distribución chi-cuadrado (gl = 2) del ejemplo acerca de la tendencia de descanso y medio de transporte, que muestra el punto de corte para rechazar la hipótesis nula al nivel 0,01.



1. **Replantear el problema en función de hipótesis de investigación e hipótesis nula de las poblaciones.** Las dos poblaciones son:

Población 1: alumnos como los entrevistados.
Población 2: alumnos cuyo abandono o continuidad en la facultad durante el primer semestre es independiente del hecho de ser o no la primera generación en la familia que asiste a la universidad.

La hipótesis nula establece que las dos poblaciones son iguales y que, en general, si los alumnos abandonan o no durante el primer semestre es independiente de que sean la primera generación de su familia que asiste a la universidad. La hipótesis de investigación establece que las poblaciones no son iguales. En otras palabras, la hipótesis de investigación establece que los alumnos como los entrevistados, no son iguales a la población hipotética en la que abandonar no está relacionado con que sean la primera generación.

2. **Determinar las características de la distribución comparativa.** Es una distribución chi-cuadrado con 1 grado de libertad.

3. **Determinar el punto de corte en la distribución comparativa, a partir del cual debería rechazarse la hipótesis nula.** Utilizando el nivel 0,01 y 1 grado de libertad, el chi-cuadrado necesario para alcanzar significación es 6,635. La figura 14-5 representa gráficamente este cálculo.

4. **Determinar el valor muestral en la distribución comparativa.** Para calcular el chi-cuadrado, primero debemos calcular las frecuencias esperadas para cada casilla. Las frecuencias mencionadas se calculan multiplicando los porcentajes esperados por la cantidad en la muestra. Por ejemplo, analicemos los abandonos de la primera generación. En general, los abandonos representan el 7,9% de los alumnos; por lo tanto, si la hipótesis nula fuera verdadera, los abandonos deberían representar el 7,9% de los 730 alumnos de la primera generación. Es decir, la frecuencia esperada para los abandonos de la primera generación es 57,7 (7,9% x 730 = 57,7). Una vez que hemos calculado las frecuencias esperadas para cada casilla, el resto del análisis chi-cuadrado sigue el procedimiento habitual que implica calcular la diferencia de cada casilla, elevarla al cuadrado, dividirla por la frecuencia esperada y sumar los resultados de todas las casillas. Como lo indica la tabla 14-8, el resultado es un chi-cuadrado de 6,73.

5. **Comparar los valores obtenidos de los pasos 3 y 4 para determinar si se rechaza o no la hipótesis nula.** El chi-cuadrado de 6,73 es mayor que el punto de corte de 6,635 (*véase* figura 14-5). Por lo tanto, la conclusión es rechazar la hipótesis nula. Es decir, a juzgar por una muestra de 2.045 alumnos de la Universidad de Indiana, los alumnos que son la primera generación de su familia en asistir a la universidad tienen más posibilidades que los demás alumnos de abandonar durante el primer semestre. (No debemos olvidar, por supuesto, que podría haber muchas razones para este resultado).

Un tercer ejemplo

Janice Steil y Jennifer Hay (1997) realizaron una encuesta a profesionales (abogados, doctores, banqueros, etc.) acerca de cuáles eran las personas con las que se comparaban cuando pensaban en su situación laboral (salario, beneficios, responsabilidades, nivel social, etc.). Una de las cuestiones de mayor interés era averiguar cuántos profesionales se comparaban a sí mismos con personas de su propio sexo, del sexo opuesto, o ambos.

La tabla 14-9 muestra los resultados junto con el porcentaje correspondiente a cada tipo de comparación, más las frecuencias esperadas (que aparecen entre paréntesis) sobre la base de esos porcentajes. Debajo de la tabla de contingencia están los cálculos de la prueba chi-cuadrado de independencia.

Tabla 14-8.
Resultados y cálculos de la prueba chi-cuadrado de independencia que prueba si la primera generación de alumnos universitarios difiere de otras en cuanto a abandono de estudios durante el primer semestre.

	Generación que asiste a la universidad		Total
	Primera	*Otras*	
Abandono	73 (57,7)	89 (103,9)	162 (7,9%)
Continuidad	657 (672,3)	1,226 (1.211,1)	1.883 (92,1%)
	730	1,315	2,045

$gl = (N_{\text{Columnas}} - 1)(N_{\text{Filas}} - 1) = (2 - 1)(2 - 1) = (1)(1) = 1$

Punto de corte chi-cuadrado necesario, $gl = 1$, nivel 0,01: 6,635

$$\chi^2 = \Sigma \frac{(O-E)^2}{E} = \frac{(73-57,7)^2}{57,7} + \frac{(89-103,9)^2}{103,9} + \frac{(657-672,3)^2}{672,3} + \frac{(1.226-1.211,1)^2}{1.211,1}$$

$$= \frac{15,3^2}{57,7} + \frac{-14,9^2}{103,9} + \frac{-15,3^2}{672,3} + \frac{14,9^2}{1.211,1}$$

$$= \frac{234,1}{57,7} + \frac{222}{103,9} + \frac{234,1}{672,3} + \frac{222}{1.211,1}$$

$$= 4,06 + 2,14 + 0,35 + 0,18$$
$$= 6,73$$

Conclusión: se rechaza la hipótesis nula.

Nota: 1. Con un análisis 2 x 2, las diferencias y las diferencias cuadráticas (numeradores) de las casillas son idénticas. *En el ejemplo que analizamos, las diferencias se deben al redondeo.* 2. **Fuente**: Riehl (1994). El chi-cuadrado exacto (6,73) es levemente diferente al informado en el artículo (7,2) debido a diferencias de redondeo.

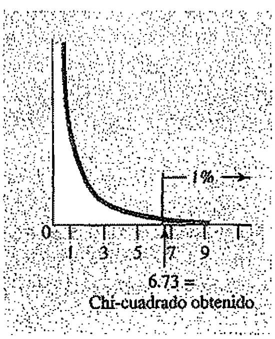


Figura 14-5.
Distribución chi-cuadrado (gl= 1) del ejemplo de Riehl (1994), que muestra el punto de corte para rechazar la hipótesis nula al nivel 0,01.

1. **Replantear el problema en función de hipótesis de investigación e hipótesis nula de las poblaciones.** Las dos poblaciones son:

Población 1: profesionales como los entrevistados.
Población 2: profesionales para quienes el propio sexo es independiente del sexo de aquellos con quienes comparan su situación laboral.

La hipótesis nula establece que las dos poblaciones son iguales, que en general los hombres y mujeres profesionales no difieren en cuanto al sexo de aquellos con quienes comparan su situación laboral. La hipótesis de investigación establece que las poblaciones no son iguales, que los profesionales como los entrevistados no son iguales a la población hipotética en la que hombres y mujeres no difieren en cuanto al sexo de aquellos con quienes comparan sus situaciones laborales.

2. **Determinar las características de la distribución comparativa.** Es una distribución chi-cuadrado con 2 grados de libertad.

3. **Determinar el punto de corte en la distribución comparativa, a partir del cual debería rechazarse la hipótesis nula.** Utilizando el nivel 0,05 y teniendo 1 grado de libertad, el chi-cuadrado necesario para la significación es 5,992. La figura 14-6 representa gráficamente este cálculo.

4. **Determinar el valor muestral en la distribución comparativa.** Como siempre, primero calculamos las frecuencias esperadas para cada casilla multiplicando los porcentajes esperados por la cantidad en la muestra. Por ejemplo, el 39% de los profesionales en general se comparan con otros del mismo sexo. Si la hipótesis nula es verdadera y las variables son independientes, se espera que el 39% de los 59 hombres se comparen con otros del mismo sexo; la frecuencia esperada de hombres que se comparan con otros del mismo sexo es 23 (es decir, 39% x 59 = 23). (Probablemente el alumno ya haya notado que en este ejemplo las frecuencias esperadas para cada tipo de comparación son las mismas entre las dos columnas. Si bien no es lo acostumbrado en una tabla de contingencia de 2 x 3, en el ejemplo resulta de ese modo porque existe la misma cantidad de personas en cada columna).

Tabla 14-9.
Resultados y cálculos de la prueba chi-cuadrado de independencia que prueba si hombres y mujeres difieren en cuanto al sexo de las personas con las que comparan su situación laboral.

		Respuesta		Total
		Hombres	*Mujeres*	
Comparación	*Mismo sexo*	29 (23)	17 (23)	46 (39,0%)
	Sexo opuesto	4 (9)	14 (9)	18 (15,3%)
	Ambos sexos	26 (27)	28 (27)	54 (45,8%)
		59	59	118

$gl = (N_{\text{Columnas}} - 1)(N_{\text{Filas}} - 1) = (2-1)(3-1) = (1)(2) = 2$

Chi-cuadrado necesario, $gl = 2$, nivel 0,05: 5,992

$$\chi^2 = \Sigma \frac{(O-E)^2}{E} = \frac{(29-23)^2}{23} + \frac{(17-23)^2}{23} + \frac{(4-9)^2}{9} + \frac{(14-9)^2}{9} + \frac{(26-27)^2}{27} + \frac{(28-27)^2}{27}$$

$$= \frac{6^2}{23} + \frac{-6^2}{23} + \frac{-5^2}{9} + \frac{5^2}{9} + \frac{-1^2}{27} + \frac{1^2}{27}$$

$$= \frac{36}{23} + \frac{36}{23} + \frac{25}{9} + \frac{25}{9} + \frac{1}{27} + \frac{1}{27}$$

$$= 1{,}57 + 1{,}57 + 2{,}78 + 2{,}78 + 0{,}04 + 0{,}04 = 8{,}78$$

Conclusión: se rechaza la hipótesis nula.

Fuente: Steil & Hay (1997). El chi-cuadrado calculado aquí (8,78) es levemente diferente del informado en la publicación (8,76) debido a diferencias de redondeo.

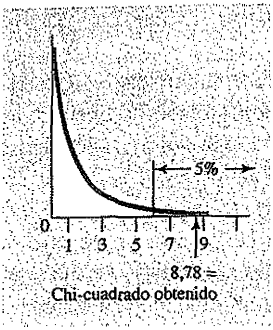


Figura 14-6.
Distribución chi-cuadrado ($gl = 2$) del ejemplo de Steil y Hay (1997), que muestra el punto de corte para rechazar la hipótesis nula al nivel 0,05.

Una vez que calculamos las frecuencias esperadas para cada casilla, el resto del análisis chi-cuadrado sigue el procedimiento habitual: calcular la diferencia para cada casilla, elevarlas al cuadrado, dividirlas por las frecuencias esperadas y sumar los resultados de todas las casillas. Tal como lo indica la tabla 14-9, el resultado es un chi-cuadrado de 8,78.

5. **Comparar los valores obtenidos en los pasos 3 y 4 para determinar si se rechaza o no la hipótesis nula.** El chi-cuadrado de 8,78 es mayor que el punto de corte de 5,992 (*véase* figura 14-6); por lo tanto, podemos rechazar la hipótesis nula. Es decir, basándonos en el ejemplo, el sexo de las personas con las que se comparan los profesionales con respecto a su situación laboral es probablemente diferente para hombres y mujeres. Al analizar las frecuencias de las casillas observadas, la mayor diferencia parece ser que las mujeres tienen aproximadamente las mismas probabilidades de compararse con otras personas del mismo sexo o del sexo opuesto, mientras que es mucho más probable que los hombres se comparen con personas del mismo sexo que con las del sexo opuesto.

SUPUESTOS DE LAS PRUEBAS CHI-CUADRADO

La prueba chi-cuadrado no requiere los supuestos usuales de normalidad de la población, de igualdad de varianzas u otros similares. Existe, sin embargo, un supuesto clave: no debe existir *ninguna relación especial de ninguno de los valores observados con algún otro valor observado*. Básicamente, lo anterior significa que no se pueden utilizar las pruebas chi-cuadrado usuales si las observaciones se basan en las mismas personas puestas a prueba más de una vez. Tomemos, por ejemplo, un estudio en el que 20 personas fueron probadas para observar si la distribución de su marca preferida de cereal para el desayuno cambió entre antes y después de una campaña reciente sobre nutrición. Los resultados de ese estudio no podrían ser probados con el chi-cuadrado usual.

TAMAÑO DEL EFECTO Y POTENCIA DE LAS PRUEBAS CHI-CUADRADO DE INDEPENDENCIA

En las pruebas chi-cuadrado de independencia podemos calcular un tamaño del efecto estimado utilizando el chi-cuadrado que calculamos. El tamaño del efecto estimado indica el grado de relación entre las dos variables nominales. En un caso extremo, dos variables nominales podrían no tener ninguna relación, es decir, ser independientes la una de la otra. En esa situación de ausencia de relación, *el tamaño del efecto estimado es cero. En otro caso extremo*, las dos variables podrían estar perfectamente relacionadas entre sí, lo cual significaría que saber en qué categoría se encuentra una persona con respecto a una variable nominal nos indicaría exactamente en qué categoría se encuentra con respecto a la otra variable nominal. (Por ejemplo, supongamos que el resultado del estudio de Riehl hubiera sido que **todos** los *alumnos* de la primera generación abandonan y que **ninguno** de los otros alumnos lo hace. En ese caso, existe una relación perfecta. Por lo tanto, si sabemos que una persona es un alumno de la primera generación sabríamos que es de los que abandonan, y si supiéramos que alguien es alumno de generaciones siguientes a la primera, sabríamos que no es de los que abandonan). Cuando existe una relación perfecta, el tamaño del efecto chi-cuadrado estimado resulta ser 1,0. En la mayoría de los casos reales, la relación de dos variables nominales se ubica entre la ausencia de relación (independiente), y una relación perfecta. Por lo tanto, la estimación del tamaño del efecto chi-cuadrado comúnmente se encuentra entre 0 y 1. Cuanto más cerca se encuentra de 0 el tamaño del efecto estimado, menor será la relación, o bien, más cerca estarán de la independencia las dos variables nominales. Cuanto más cerca esté de 1 el tamaño del efecto estimado, mayor será la relación o más cerca estarán las dos variables nominales de una relación perfecta.[3]

En una tabla de contingencia 2 x 2, la medida de asociación se denomina **coeficiente phi** (ϕ). Es la raíz cuadrada del resultado de la división del chi-cuadrado por la cantidad de personas en toda la muestra. Se expresa bajo la fórmula:

$$\phi = \sqrt{\frac{\chi^2}{N}} \tag{14-4}$$

Las reglas de Cohen (1988) para el coeficiente phi establecen que 0,10 es un tamaño del efecto pequeño, 0,30 es un tamaño del efecto mediano y 0,50 es un gran tamaño del efecto.

Por ejemplo, en el estudio de Riehl acerca de la primera generación de estudiantes universitarios, el chi-cuadrado que calculamos era de 6,7, y había 2.045 personas en el estudio. Aplicando la fórmula para el coeficiente phi,

$$\phi = \sqrt{\frac{\chi^2}{N}} = \sqrt{\frac{6{,}7}{2.045}} = \sqrt{0{,}00328} = 0{,}06$$

Se trata de un tamaño del efecto muy pequeño. Los resultados con respecto a la significación nos indican que la mayor probabilidad de que los alumnos de primera generación abandonen los estudios, probablemente no es casual. Pero el coeficiente phi nos indica que, en la práctica, esa diferencia no casual no puede ser un factor muy importante. (En el capítulo 8 tratamos aquellas situaciones en la que un resultado es estadísticamente significativo pero cuyo tamaño de efecto es muy pequeño).

El estadístico phi sólo se aplica cuando existe una situación 2 x 2. Podemos decir que el estadístico **phi de Cramer** es una extensión del coeficiente phi ordinario, que puede aplicarse a tablas de contingencia mayores de 2 x 2. (El phi de Cramer también se conoce como la *V* de Cramer, y a veces se escribe ϕ_C ó V_C). Se calcula del mismo modo que el coeficiente phi ordina-

[3] Si el alumno ya ha estudiado el capítulo 3, puede resultarle útil considerar un tamaño del efecto chi-cuadrado estimada como un coeficiente de correlación. De hecho, en el caso de una tabla de contingencia 2 x 2, la estimación es realmente idéntica al coeficiente de correlación. Supongamos que tomáramos las dos variables de una tabla de contingencia 2 x 2 y arbitrariamente hiciéramos que uno de los valores de cada uno fuera 1 y el otro fuera igual a 0. Si después calculáramos un coeficiente de correlación entre las dos variables, el resultado sería exactamente el mismo que el coeficiente phi descripto en el siguiente párrafo (no obstante, según qué categorías de cada variable hayamos transformado en 1 ó en 0, la correlación será negativa o positiva). Las reglas de Cohen de tamaños del efecto pequeños, medianos y grandes para el coeficiente phi, que se describen a continuación, también son exactamente las mismas que las de un coeficiente de correlación.

rio, excepto que en lugar de dividir por N, se divide por el resultado de la multiplicación de N por los grados de libertad del lado menor de la tabla (gl_{Menor}). Se expresa bajo la fórmula,

$$\phi \text{ de Cramer} = \sqrt{\frac{\chi^2}{(N)(gl_{\text{Menor}})}} \qquad (14\text{-}5)$$

En el ejemplo acerca de las preferencias de transporte, el chi-cuadrado era de 16,07 y la cantidad total de personas entrevistadas era 200. Los grados de libertad del lado más pequeño de la tabla (en este caso las filas) era 1. El phi de Cramer es la raíz cuadrada del resultado de dividir 16,07 por 200 por 1, es decir, 0,28. Lo anterior se expresa bajo la fórmula,

$$\phi \text{ de Cramer} = \sqrt{\frac{\chi^2}{(N)(gl_{\text{Menor}})}} = \sqrt{\frac{16{,}07}{(200)(1)}} = \sqrt{0{,}08} = 0{,}28$$

En el estudio de Steil y Hay acerca del sexo con el que hombres y mujeres profesionales se comparaban a sí mismos, calculamos un chi-cuadrado de 8,78, y se entrevistó a 118 profesionales. Los grados de libertad del lado más pequeño de la tabla (en este caso las columnas) era 1. El phi de Cramer es 0,27 (la raíz cuadrada de 8,78 dividido 118 es 0,27). Lo anterior se expresa bajo la fórmula:

$$\phi \text{ de Cramer} = \sqrt{\frac{\chi^2}{(N)(gl_{\text{Menor}})}} = \sqrt{\frac{8{,}78}{(118)(1)}} = \sqrt{0{,}07} = 0{,}26$$

Las reglas de Cohen para el tamaño del efecto del phi de Cramer dependen de los grados de libertad del lado menor de la tabla. La tabla 14-10 muestra las reglas de Cohen para el tamaño del efecto del phi de Cramer (1988), correspondientes a tablas cuyo lado menor es 2, 3 y 4. Cabe destacar que cuando el lado menor de la tabla es 2, el grado de libertad es 1 y, por lo tanto, los tamaños del efecto que indica la tabla para esa situación son los mismos que para el coeficiente phi ordinario. (Dado que multiplicar por 1 no produce ningún cambio, el cálculo también arroja el mismo resultado, tal como sucede en los dos ejemplos que analizamos anteriormente).

Basándonos en la tabla, en el ejemplo del transporte existe un tamaño del efecto aproximadamente mediano (0,28), es decir, una relación mediana entre el tipo de transporte utilizado y el hecho de que se trate de una persona diurna o nocturna.

Tabla 14-10.
Reglas de Cohen para el phi de Cramer.

Menor dimensión de la tabla de contingencia	Tamaño del efecto		
	Pequeño	*Mediano*	*Grande*
2 ($gl_{\text{Menor}} = 1$)	0,10	0,30	0,50
3 ($gl_{\text{Menor}} = 2$)	0,07	0,21	0,35
4 ($gl_{\text{Menor}} = 3$)	0,06	0,17	0,29

Potencia y tamaño de muestra necesarios para la prueba chi-cuadrado de independencia

La tabla 14-11 muestra la potencia aproximada al nivel 0,05 de significación, para tamaños del efecto pequeños, medianos y grandes y tamaños totales de muestra de 25, 50, 100 y 200. Se indica la potencia para tablas con 1, 2, 3 y 4 grados de libertad.[4]

Por ejemplo, analicemos la potencia de un estudio planificado de 2 x 4 ($gl = 3$) con 50 personas, con un tamaño del efecto esperado mediano (phi de Cramer = 0,30), que se realizará con un nivel de 0,05. Utilizando la tabla 14-11, el estudio mencionado tendría una potencia de 0,40, es decir, que si la hipótesis de investigación en realidad es verdadera y el tamaño del efecto real es mediano, existe aproximadamente un 40% de posibilidades de que el estudio resulte significativo.

La tabla 14-12 indica la cantidad total aproximada de participantes necesarios para obtener una potencia del 80%, con tamaños del efecto pequeños, medianos y grandes, a un nivel 0,05 de significación, para pruebas chi-cuadrado de independencia con 2, 3, 4 y 5 grados de libertad.[5] Por

Tabla 14-11.
Potencia aproximada para una prueba chi-cuadrado de independencia en la que se prueba la hipótesis al nivel 0,05 de significación.

gl Total	*N* Total	Tamaño del efecto		
		Pequeño ($\phi = 0{,}10$)	*Mediano* ($\phi = 0{,}30$)	*Grande* ($\phi = 0{,}50$)
1	25	0,08	0,32	0,70
	50	0,11	0,56	0,94
	100	0,17	0,85	*
	200	0,29	0,99	*
2	25	0,07	0,25	0,60
	50	0,09	0,46	0,90
	100	0,13	0,77	*
	200	0,23	0,97	*
3	25	0,07	0,21	0,54
	50	0,08	0,40	0,86
	100	0,12	0,71	0,99
	200	0,19	0,96	*
4	25	0,06	0,19	0,50
	50	0,08	0,36	0,82
	100	0,11	0,66	0,99
	200	0,17	0,94	*

*Casi 1.

[4] Cohen (1988, pp. 228-248) proporciona tablas más detalladas. Las tablas de Cohen se basan en un tamaño del efecto denominado *w*, que es equivalente al phi pero no al phi de Cramer. En la página 222, Cohen ofrece también una útil tabla de conversión de phi de Cramer a *w*.

[5] Cohen (1988, pp. 253-267) proporciona tablas más detalladas. Para utilizar esas tablas, debe tenerse en cuenta lo indicado en la nota al pie nº 4. Además, Dunlap y Myers (1997) han demostrado que, con una tabla 2 x 2, la cantidad aproximada de participantes necesarios para una potencia de 80 – 90% es $8/\phi^2$.

Tabla 14-12.
Cantidad total aproximada de participantes necesarios para una potencia del 80% en una prueba chi-cuadrado de independencia, en la que se prueba la hipótesis al nivel 0,05 de significación.

gl Total	Tamaño del efecto		
	Pequeño ($\phi = 0{,}10$)	*Mediano* ($\phi = 0{,}30$)	*Grande* ($\phi = 0{,}50$)
1	785	87	26
2	964	107	39
3	1.090	121	44
4	1.194	133	48

ejemplo, supongamos que planificamos un estudio con una tabla de contingencia 3 x 3 ($gl = 4$), que esperamos un gran tamaño del efecto y que utilizamos el nivel 0,05 de significación. De acuerdo con la tabla, sólo necesitaríamos 48 participantes (aproximadamente 5 ó 6 por casilla).

CONTROVERSIAS Y LIMITACIONES

Hace medio siglo, Lewis y Burke (1949) publicaron un trabajo memorable acerca de la utilización inadecuada del chi-cuadrado. Enumeraron nueve errores comunes aparecidos en publicaciones y dieron numerosos ejemplos de cada uno de ellos. Con una sola excepción, su obra sigue vigente a través de los años. Los errores siguen cometiéndose, y aún siguen considerándoselos errores.

La única excepción de esa descripción crítica es el error que Lewis y Burke consideraban la debilidad más común en la utilización del chi-cuadrado: frecuencias esperadas demasiado bajas. En la actualidad, aparentemente esperar cantidades pequeñas para las casillas puede no ser un problema tan grave. Lewis y Burke, como la mayoría de los autores de textos sobre estadística de su tiempo, sostenían que cada casilla de una tabla de contingencia (y cada categoría de una prueba de bondad de ajuste) debería tener una frecuencia esperada de tamaño razonable. Recomendaban un mínimo de 10, siendo 5 la cantidad límite inferior. Otros recomendaban cifras que iban del 1 al 20. Incluso Sir Ronald Fisher (1938) tomó partido, recomendando 10 como mínimo. Asimismo, otros recomendaban que el mínimo debía ser una proporción del total, o que dependía del hecho de que las frecuencias esperadas fueran iguales o no. (A propósito, cabe mencionar que lo que se estaba debatiendo eran frecuencias mínimas **esperadas**, no frecuencias observadas)

Desde el año 1949, cuando Lewis y Burke publicaron su trabajo, han habido algunas investigaciones sistemáticas acerca de cuáles eran exactamente los efectos de pequeñas frecuencias esperadas. (En esos estudios se aplican los métodos de Montecarlo; *véase* cuadro 10-1). ¿Cuál es la conclusión? Al igual que en la mayoría de las áreas, la controversia aún no está totalmente definida. Sin embargo, una importante revisión de las investigaciones realizadas sobre el tema (Delucchi, 1983) plantea dos conclusiones principales:

1. "Como norma general, el chi-cuadrado puede usarse apropiadamente en casos en que los valores esperados son mucho menores a los que anteriormente se consideraban permisibles" (p. 168). Incluso frecuencias esperadas tan bajas como 1 por casilla pueden ser aceptables en términos del error Tipo I, siempre que en total exista una cantidad razonable de individuos.

Aparentemente, el principio más importante es que la cantidad de individuos debería ser, al menos, cinco veces la cantidad de casillas. Por ejemplo, una pequeña frecuencia esperada sería aceptable en una tabla de contingencia 2 x 2, si hubiera al menos 20 participantes en el estudio.[6]

2. Sin embargo, Delucchi cita a un investigador que concluye que, aunque puede ser aceptable utilizar el chi-cuadrado con pequeñas frecuencias esperadas (en el sentido de que a la larga no produce demasiados errores Tipo I), de todos modos no es un método muy sensato, ya que la posibilidad de obtener un resultado significativo, aun si la hipótesis de investigación es verdadera, puede ser bastante escaso. Es decir, con pequeñas frecuencias esperadas, la potencia es muy baja y entonces se corre el riesgo de que se produzcan errores Tipo II.

LAS PRUEBAS CHI-CUADRADO SEGÚN SE DESCRIBEN EN LAS PUBLICACIONES CIENTÍFICAS

Los informes de pruebas chi-cuadrado generalmente incluyen las frecuencias en cada categoría o casilla, al igual que los grados de libertad, la cantidad de participantes, el chi-cuadrado calculado y el nivel de significación. Por ejemplo, Harter et al. informaron sus hallazgos con respecto al estilo de relación de los hombres concentrados en sí mismos como "$\chi^2(2, n = 101) = 11{,}89, p < 0{,}005$" (p. 156).

Veamos otro ejemplo completo de prueba chi-cuadrado de bondad de ajuste. Sandra Moriarty y Shu-Ling Everett (1994) realizaron un estudio acerca del modo de mirar televisión. En el marco del estudio, un grupo de alumnos de posgrado visitaron 55 hogares diferentes y observaron a los televidentes durante sesiones de 45 minutos. Una parte de los resultados comparaba qué cantidad de personas observadas podían ser incluidas en una de cuatro categorías diferentes:

> El *zapping* (cambiar de canal muy rápidamente), la categoría dominada por el tipo de comportamiento más activo, ocurrió con más frecuencia en el 33% de las sesiones ($n = 18$). La categoría pastoreo (curiosear los canales durante algunos períodos) dominó el 24% de las sesiones ($n= 13$), y un 22% correspondía a cada una de las categorías de visión continua y prolongada ($n = 12$). Las diferencias no fueron estadísticamente significativas ($\chi^2 = 1{,}79, gl = 3, p > 0{,}05$).

Las publicaciones de pruebas chi-cuadrado de independencia proporcionan la misma información básica acerca de los chi-cuadrados. Por ejemplo, los resultados del estudio de Steil y Hay sobre los profesionales se informaron del siguiente modo: "χ^2 (2, $N = 118$) = 8,76, $p < 0{,}05$" (p. 432).

Veamos otro ejemplo "tomado de una publicación" acerca del modo en que se informa una prueba chi-cuadrado de independencia. John Lydon y sus socios (1997) realizaron un estudio que comparaba relaciones a larga distancia y locales. Los investigadores primero repartieron cuestionarios a un grupo de alumnos un mes antes de que dejaran sus casas para comenzar su primer se-

[6] Supongamos que tenemos una tabla mayor a 2 x 2, con una categoría o casilla que tiene una frecuencia esperada extremadamente pequeña (o incluso una frecuencia esperada moderadamente pequeña si la cantidad de participantes también es pequeña). Una solución es combinar categorías relacionadas para aumentar la frecuencia esperada y reducir la cantidad total de casillas. Sin embargo, la anterior es una solución de último recurso si la adaptación se realiza basándose en los resultados del experimento. El problema es que se estaría capitalizando el hecho de conocer el resultado. La mejor solución es agregar personas al estudio, pero si esto no fuera factible, a veces se puede aplicar un procedimiento alternativo, denominado "prueba exacta de Fisher", que se describe en algunos textos sobre estadística de nivel intermedio.

mestre en la Universidad Mc Gill (tiempo 1). Algunos de estos alumnos tenían parejas que vivían en el área de McGill, otros tenían parejas que vivían lejos de Mc Gill. Los investigadores se pusieron en contacto con los participantes nuevamente durante el semestre de otoño, preguntándoles por el estado actual de sus relaciones de pareja (tiempo 2). El siguiente es el informe de sus resultados:

> De los 69 participantes [...] 55 estaban involucrados en relaciones a larga distancia y 14 en relaciones locales (parejas que vivían dentro de los 200 km de donde vivían ellos). Coherentemente con nuestras predicciones, 12 de las 14 relaciones locales estaban intactas al tiempo 2 (86%), mientras que sólo 28 de las 55 relaciones a distancia permanecían intactas (51%). $\chi^2(1, N = 69) = 5{,}55$, $p < 0{,}02$. (p. 108)

Aunque Lydon et al. no indicaron el tamaño del efecto del resultado significativo, podemos calcularlo a partir de la información proporcionada. El cálculo estadístico apropiado para el tamaño del efecto es el coeficiente phi, ya que se trata de una tabla chi-cuadrado 2 x 2 (local contra larga distancia x intacto contra terminada). Si aplicamos la fórmula:

$$\phi = \sqrt{\frac{\chi^2}{N}} = \sqrt{\frac{5{,}55}{69}} = \sqrt{0{,}08} = 0{,}028$$

El resultado de la fórmula sugiere que existe un tamaño del efecto moderado.

Resumen

Las pruebas chi-cuadrado son pruebas de hipótesis para variables nominales. El chi-cuadrado mide el grado de discrepancia entre frecuencias esperadas y observadas de varios niveles o categorías. Se calcula encontrando la diferencia entre la frecuencia observada y la frecuencia esperada de cada categoría o combinación de categorías, elevando esa diferencia al cuadrado (para eliminar signos positivos y negativos) y dividiéndola por la frecuencia esperada (para que las diferencias cuadráticas sean más proporcionales a las cantidades involucradas). Luego se suman los resultados de todas las categorías o combinaciones de categorías. La distribución chi-cuadrado es una distribución conocida, y los puntos de cortes pueden encontrarse en una tabla estándar.

La prueba chi-cuadrado de bondad de ajuste se utiliza para probar la hipótesis de que una distribución de frecuencias de dos o más categorías de una variable nominal coincide con una distribución esperada. (Las frecuencias esperadas se basan, por ejemplo, en una teoría o en una distribución de otro estudio o circunstancia). En este tipo de pruebas, las frecuencias esperadas se indican de antemano o se basan en algunos porcentajes esperados (como por ejemplo, el mismo porcentaje para todos los grupos). Los grados de libertad son la cantidad de categorías menos 1.

La prueba chi-cuadrado de independencia se utiliza para probar la hipótesis sobre la relación entre dos variables nominales, es decir, si el esquema de repetición de los participantes en la categoría de una variable tiene el mismo patrón proporcional dentro de cada una de las categorías de la otra variable. Los datos se exponen en una tabla de contingencia, en la que las dos variables se cruzan y las cantidades de participantes de cada combinación se ubican dentro de cada una de las casillas resultantes. La frecuencia esperada para una casilla, si las dos variables son independientes, es el porcentaje de todas las personas en la fila de la casilla multiplicado por la cantidad total de personas en la columna de esa casilla. Los grados de libertad para la prueba de independencia son la cantidad de columnas menos 1, multiplicada por la cantidad de filas menos 1.

El tamaño del efecto estimado para una prueba chi-cuadrado de independencia (el grado de asociación), con una tabla de contingencia 2 x 2, es el coeficiente phi; y con tablas mayores, es el phi de Cramer. Phi es la raíz cuadrada del resultado de la división del chi-cuadrado calculado por la cantidad de participantes. El phi de Cramer es la raíz cuadrada del resultado de la división del chi-cuadrado, calculado por el producto de la cantidad de participantes por los grados de libertad de la dimensión más pequeña de la tabla de contingencia. Estos coeficientes van de 0 a 1; 0 indica una independencia perfecta y 1 una relación perfecta. Un phi de 0,10 se considera un tamaño del efecto pequeño, de 0,30 un tamaño del efecto mediano y de 0,50 un gran tamaño del efecto.

Las pruebas chi-cuadrado no tienen supuestos relacionados con las distribuciones normales de sus variables, pero sí requieren que la categoría o casilla en la que se incluye a un participante sea independiente de la categoría o casilla de cualquier otro participante.

La frecuencia mínima aceptable para una categoría o casilla ha sido tema de controversias. Actualmente, el mejor consejo es tener en cuenta que, incluso pequeñas frecuencias esperadas, no aumentan seriamente las posibilidades de un error Tipo I, siempre que haya al menos una cantidad de individuos igual a cinco veces la cantidad de categorías (casillas).

No obstante, las pequeñas frecuencias esperadas reducen seriamente la potencia y deben evitarse siempre que sea posible.

Términos clave

- Variable categórica.
- Distribución chi-cuadrado.
- Chi-cuadrado (χ^2).
- Tabla chi-cuadrado.
- Prueba chi-cuadrado de bondad de ajuste.
- Prueba chi-cuadrado de independencia.
- Tabla de contingencia.
- Phi de Cramer.
- Frecuencia esperada.
- Independencia.
- Variable nominal.
- Frecuencia observada.
- Coeficiente phi (ϕ).

Ejercicios

Los ejercicios implican la realización de cálculos (con la ayuda de una calculadora). La mayoría de los problemas estadísticos reales se resuelven por computadora, pero aunque exista la posibilidad de utilizarla, es conveniente realizar estos ejercicios manualmente para incorporar el método de trabajo.

Para adquirir práctica en la utilización de una computadora, para resolver problemas estadísticos, se puede utilizar la sección de computación de cada capítulo, publicada en la *Guía de estudio y libro de tareas de computación para el alumno [Student's Study Guide and Computer Workbook]* que acompaña este libro.

Todos los datos de esta sección son ficticios (a menos que se especifique lo contrario).

Las respuestas a los ejercicios de la serie I se encuentran al final del libro.

SERIE I

1. Calcule una prueba chi-cuadrado de bondad de ajuste para cada uno de los siguientes casos (utilice el nivel 0,05 para cada uno):

(a)	**Categoría**	**Esperado**	**Observado**
	A	20%	19
	B	20%	11
	C	40%	10
	D	10%	15
	E	10%	15
(b)	**Categoría**	**Esperado**	**Observado**
	I	30%	100
	II	50%	100
	III	20%	100

(c)

Categoría	Cantidad en el pasado	Observado
1	100	38
2	300	124
3	50	22
4	50	16

(d)

Categoría	Observado[a]
Artes	37
Ciencias	21
Humanidades	32

[a] Lo esperado es la misma cantidad en cada categoría.

2. Un director de una pequeña clínica de psicoterapia intenta planificar la contratación de personal temporario para que colabore con la tarea de admisión de los pacientes, y se pregunta si la actividad de la clínica difiere entre las distintas temporadas del año. El último año ingresaron 28 nuevos pacientes en invierno, 33 en primavera, 16 en verano y 51 en el otoño. Sobre la base de la información del año anterior, ¿el director debería concluir que existe una diferencia entre las distintas estaciones? (Utilice un nivel de 0,05) a) Realice los cinco pasos de la prueba de hipótesis. b) Explique su respuesta a una persona que nunca ha tomado un curso de estadística. (**Nota**: este ejercicio es similar al ejemplo de Harter et al., en el que se calcula un chi-cuadrado para una sola variable nominal. No se trata de una prueba chi-cuadrado de independencia y no incluye tablas de contingencia).

3. Folwell et al. (1997) entrevistaron a un grupo de adultos, de 54 años o mayores, acerca de la relación que mantenían con sus hermanos. Una de las preguntas planteaba si había ocurrido algún cambio en la proximidad emocional a través de los años. Descubrieron que 43 de los entrevistados "percibieron cambios en la proximidad emocional de las relaciones con sus hermanos [...] [y] 14 no informaron cambios en la proximidad de la relación con sus hermanos" (p. 846). También probaron si la diferencia era mayor de lo que se esperaría por casualidad (que sería una proporción de 50 y 50). "Un análisis chi-cuadrado reveló que los entrevistados percibieron cambios en la proximidad de las relaciones con sus hermanos ($\chi^2 = 14{,}75$, $gl = 1$, $\alpha = 0{,}05$)". (p. 846)

a) Calcule usted mismo el chi-cuadrado (muestre su trabajo). Sus resultados deberían ser iguales a los anteriores, teniendo en cuenta las diferencias por redondeo. b) Explique el resultado a una persona que nunca ha tomado un curso de estadística.

4. Realice una prueba chi-cuadrado de independencia para cada una de las siguientes tablas de contingencia (utilice el nivel 0,01). Además, calcule el phi o phi de Cramer para cada una:

(a)

10	16
16	10

(b)

100	106
106	100

(c)

100	160
160	100

(d)

10	16	10
16	10	10

(e)

10	16	16
16	10	16

(f)

10	16	10
16	10	16

5. Un psicólogo especializado en educación está interesado en saber si los alumnos que utilizan máquinas de escribir o procesadores de texto (o ninguno de ellos) para escribir, cuando realizan tareas en sus hogares, tienden a utilizar lapicera o lápiz cuando toman apuntes en clase. El investigador entrevista a 200 alumnos. Los resultados aparecen en la tabla que sigue a continuación. ¿Existe una relación significativa entre estas dos variables? (Utilice el nivel 0,05). a) Realice los cinco pasos de la prueba de hipótesis. b) Calcule el phi de Cramer. c) Explique su respuesta a una persona que nunca ha tomado un curso de estadística.

Elemento utilizado para tomar apuntes	Artefacto utilizado en sus hogares: *Máquina de escribir*	*Procesador de textos*	*Ninguno*
Lapicera	42	62	26
Lápiz	18	38	14

6. Un analista político está interesado en saber si existe relación entre la comunidad en la que vive una persona y la opinión de esa persona con respecto a una futura iniciativa de vo-

tación acerca de la conservación del agua. El analista entrevista a 90 personas telefónicamente y obtiene los resultados que aparecen a continuación. ¿A un nivel de 0,05, está relacionada la opinión con la comunidad? a) Realice los cinco pasos de la prueba de hipótesis. b) Calcule el phi de Cramer y la potencia. c) Explique su respuesta a una persona que nunca ha tomado un curso de estadística.

	Comunidad A	Comunidad B	Comunidad C
A favor	12	6	3
En contra	18	3	15
No emite opinión	12	9	12

7. Shi et al. (1997) entrevistaron a 558 médicos que realizan atención primaria en Carolina del Sur, concentrándose en las diferencias entre aquellos que han tratado o no a pacientes con HIV/SIDA. Algunos de los resultados aparecen en la tabla 14-13. (La tabla indica porcentajes, no números reales). Concéntrese en el resultado que aparece cerca del final y que muestra la relación entre la cantidad de horas que el médico generalmente ejerce por semana y el hecho de que el médico hubiera tratado o no a pacientes con HIV/SIDA; el resultado tiene un chi-cuadrado de 15,1. a) Calcule usted mismo el chi-cuadrado (muestre su trabajo); sus resultados deberían ser iguales a los indicados, teniendo en cuenta las diferencias de redondeo. (Para resolver el ejercicio deberá convertir los porcentajes en números reales. Observe que los porcentajes dados son los porcentajes de las personas en la columna. Por lo tanto, en el caso de los médicos que ejercen 40 horas por semana o menos y han tratado a pacientes con HIV/SIDA, el 11% es igual a 37 médicos. (Es decir, 11% x 335 = 37, redondeando para obtener la cantidad entera de médicos más cercana). b) Calcule el phi de Cramer. c) Explique su resultado a una persona que nunca ha tomado un curso en estadística.

SERIE II

1. Calcule una prueba chi-cuadrado de bondad de ajuste para cada uno de los siguientes casos (utilice el nivel 0,01 para cada uno de ellos):

(a)	Categoría	Esperado	Observado
	1	2%	5
	2	14%	15
	3	34%	90
	4	34%	120
	5	14%	50
	6	2%	20

(b)	Categoría	Proporción esperada	Observado
	A	1/3	10
	B	1/6	10
	C	1/2	10

2. Calcule una prueba chi-cuadrado para cada uno de los siguientes casos utilizando el nivel 0,05. En cada ejercicio, la distribución esperada es que las frecuencias sean iguales en todas las categorías. (Se trata de ejercicios similares al ejemplo de Harter et al. en los que realizamos un chi-cuadrado para una sola variable nominal. No son pruebas chi-cuadrado de independencia y no incluyen tablas de contingencia).

a) 5 10 5 b) 10 15 10 c) 10 20 10 d) 5 15 5

3. Una investigadora necesita estar segura de que la muestra para su estudio no deje de ser representativa de la distribución de grupos étnicos de su comunidad. La muestra incluye 300 blancos, 80 africanos americanos, 100 latinos, 40 asiáticos y 80 personas de otros grupos étnicos. En la comunidad, según registros de censos, hay un 48% de blancos, un 12% de africanos americanos, un 18% de latinos, un 9% de asiáticos y un 13% de otros grupos étnicos. ¿Es la muestra representativa o no de la población de la comunidad? (Utilice el nivel 0,05). a) Realice los cinco pasos de la prueba de hipótesis. b) Explique su respuesta a una persona que nunca ha tomado un curso de estadística. (**Nota**: el ejercicio es similar al ejemplo de Harter et al., en el que realizamos un chi-cuadrado pa-

Tabla 14-13.
Características demográficas de los médicos entrevistados (en porcentajes).

Características	Total (*N* = 558)	Han tratado pacientes con HIV/SIDA (*n* = 335)	No han tratado pacientes con HIV/SIDA (*n* = 223)	χ^2	Estadístico *t*
Especialidad de atención primaria					
Práctica familiar	34	36	31	53,8***	
Medicina interna	18	21	14		
Enfermedades infecciosas	3	5	0		
Pediatría	11	7	11		
Obstetricia/ginecología	14	8	22		
Otras	21	23	17		
Edad					
≤ 35	16	20	10	34,1***	
35-44	41	44	37		
45-54	21	21	21		
55-64	15	12	18		
65 ó más	8	3	14		
Sexo					
Masculino	89	88	89	0,7	
Femenino	11	12	11		
Origen étnico					
Minorías	9	9	8	0,1	
Raza blanca	91	91	92		
Horas por semana en que ejerce medicina clínica					
≤ 40 horas	16	11	23	15,1***	
40–49 horas	21	21	21		
≥ 50 horas	63	68	56		
Voluntad de atender					
Pacientes con HIV/SIDA[a]	3,1	3,4	2,7		5,3***

Nota: Es posible que los porcentajes no sumen 100 debido al redondeo. La mayoría de los entrevistados (51%) brindaron servicios a menos de 10 pacientes con HIV/SIDA, un 4% a 10-19 pacientes, un 1% a 20-29 pacientes, un 1% a 30-39 pacientes y un 3% a 40 ó más pacientes.

[a]La variable se midió en una escala de 5 puntos que iba del 1 = nunca a 5 = siempre.

$^{***}p < 0{,}01$.

Fuente: Shi, L., et al. (1997), tab. 1. "Médicos de atención primaria y barreras contra la atención a personas con HIV/SIDA". *Evaluación & Profesiones relacionadas con la salud [Evaluation & The Health Professions]*, 20, 164-187. Copyright © 1997, por Sage Publications, Inc. Reimpreso con autorización de Sage Publications.

ra una sola variable nominal. No es una prueba chi-cuadrado de independencia y no incluye tablas de contingencia).

4. Realice una prueba chi-cuadrado de independencia para cada una de las siguientes tablas de contingencia (utilice el nivel 0,05 y además calcule el phi o phi de Cramer y la potencia de cada una).

(a)

8	8
8	16

(b)

8	8
8	32

(c)

8	8
8	48

(d)

8	8	8
8	8	8
8	8	16

(e)

8	8	8
8	8	8
8	8	32

(f)

8	8	8
8	8	8
8	8	48

5. La siguiente tabla muestra los resultados de una encuesta realizada a una muestra de

personas que asisten al ballet, distribuidas según el tipo de ubicación que adquirieron y según la regularidad con la que asisten al ballet. ¿Existe una relación significativa? (Utilice el nivel 0,05). a) Realice los cinco pasos de la prueba de hipótesis. b) Calcule el phi de Cramer. c) Explique su respuesta a una persona que nunca ha tomado cursos de estadística.

		Asistencia	
		Regular	*Ocasional*
Categoría de ubicación	*Platea*	20	80
	Galería principal de palcos	20	20
	Galería	40	80

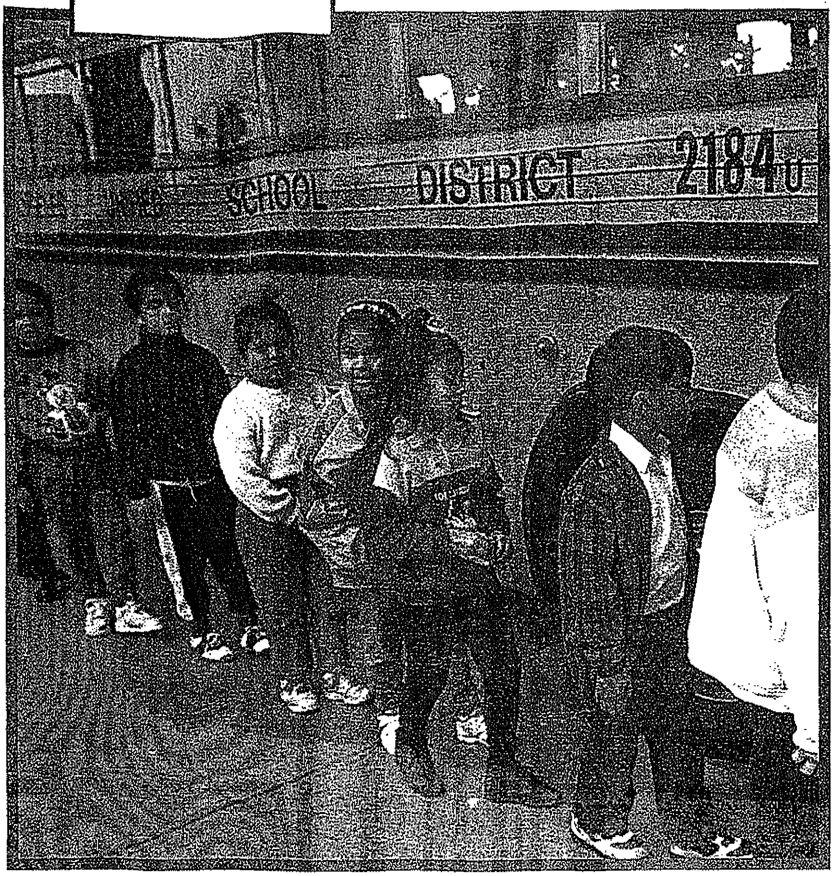
6. Everett et al. (1997) realizaron una encuesta por correo a una muestra de médicos elegidos al azar. A la mitad se les ofreció \$1 si enviaban el cuestionario (este era el grupo experimental); a la otra mitad se la utilizó como grupo control. El objetivo del estudio era comprobar si aun un pequeño incentivo aumentaría el porcentaje de devoluciones de encuestas enviadas a médicos. Everett et al. informan sus resultados de la siguiente forma:

> De las 300 encuestas enviadas por correo al grupo experimental, 39 no fueron entregadas; 2 fueron devueltas sin completar, y 164 fueron devueltas completas. Por lo tanto, el porcentaje de respuesta del grupo experimental fue del 63% (164/300-39 = 0,63). De las 300 encuestas enviadas por correo al grupo de control, 40 no fueron entregadas y 118 fueron devueltas incompletas. Por lo tanto, el porcentaje de respuestas del grupo de control fue del 45% (118/300 – 40 = 0,45). A través de una prueba chi-cuadrado que comparaba los porcentajes de respuestas de los grupos experimental y de control descubrió que el incentivo de \$1 había mejorado de forma estadísticamente significativa el porcentaje de respuesta en comparación con el grupo de control [$\chi^2(1, N = 521) = 16{,}0$, $p < 0{,}001$].

a) Calcule usted mismo el chi-cuadrado y muestre su trabajo. Sus resultados deberían ser iguales a los indicados en el párrafo anterior, teniendo en cuenta las diferencias de redondeo. (Cuando resuelva este ejercicio no olvide que se basa sólo en las respuestas que no fueron devueltas sin entregar. Por lo tanto, en el caso del grupo experimental hubo un total de 261 respuestas, de las cuales 164 fueron devueltas completas y las restantes 96 no). b) Calcule phi. c) Explique el resultado a una persona que nunca ha tomado un curso de estadística.

15 Estrategias a aplicar cuando las distribuciones poblacionales no son normales:

Transformaciones de datos, pruebas de rango y orden y métodos intensivos por computadora



Descripción del capítulo

- Supuestos de los procedimientos estándar de prueba de hipótesis.
- Transformaciones de datos.
- Pruebas de rango y orden.
- Métodos intensivos por computadora.
- Comparación de métodos.
- Controversias.
- Procedimientos que se utilizan cuando las poblaciones parecen no normales, según se describen en las publicaciones científicas.
- Resumen.
- Términos clave.
- Ejercicios.

Este capítulo analiza los procedimientos de prueba de hipótesis cuando no podemos suponer que las distribuciones poblacionales son siquiera aproximadamente normales. Al mismo tiempo, analizamos situaciones en las que no podemos cumplir con otros requisitos de los procedimientos ordinarios de prueba de hipótesis, como por ejemplo cuando las poblaciones no tienen las mismas varianzas. Primero, revisaremos brevemente la función que cumplen los supuestos en los procedimientos estándar de prueba de hipótesis. Luego, analizamos tres métodos utilizados por investigadores psicológicos cuando los resultados de un estudio no cumplen con los supuestos usuales: transformaciones de datos, pruebas de rango y orden y métodos intensivos por computadora.

SUPUESTOS DE LOS PROCEDIMIENTOS ESTÁNDAR DE PRUEBA DE HIPÓTESIS

Como vimos en los capítulos 9 al 13, para realizar una prueba t o un análisis de varianza es necesario que se cumplan ciertos supuestos. En los procedimientos de prueba de hipótesis mencionados, trabajamos con las observaciones de un estudio como si pertenecieran a poblaciones mayores, aunque desconocidas. Uno de los aspectos supuestos es que las poblaciones involucradas tienen una distribución normal; el otro supuesto principal es que las poblaciones tienen la misma varianza.

Si el alumno ya ha estudiado el capítulo 3, en lo que se refiere al coeficiente de correlación recordará que la correlación es un procedimiento descriptivo. Sin embargo, en muchas investigaciones, el coeficiente de correlación se calcula utilizando valores de una muestra de manera que el investigador pueda realizar generalizaciones acerca de una población. Cuando el propósito es el descripto, el coeficiente de correlación es más preciso si las variables que se correlacionan provienen de poblaciones distribuidas normalmente. Además, se puede calcular una prueba de significación del coeficiente de correlación (*véase* apéndice II del capítulo 3) y, también en ese caso, es muy importante el supuesto de normalidad.[1]

En capítulos anteriores también aprendimos que se obtienen resultados exactos incluso cuando un estudio sugiere que las poblaciones cumplen con los supuestos de curva normal y de igual varianza en forma muy imprecisa. Sin embargo, el tema que nos preocupa en este caso son las situaciones en las que queda claro que las poblaciones no son ni siquiera cercanas a lo normal, o que ni siquiera están cerca de tener iguales varianzas. Si en esas situaciones se utiliza la prueba t o el análisis de varianza ordinarios, se pueden obtener resultados muy incorrectos. Por ejemplo, podríamos realizar todos los cálculos correctamente y decidir rechazar la hipótesis nula conforme a esos resultados. Sin embargo, si las poblaciones no cumplen con los supuestos estándar, el resultado podría ser incorrecto (incorrecto en cuanto a que en lugar de existir realmente un 5% de probabilidad de obtener los resultados si la hipótesis nula es verdadera, ¡en realidad podría haber un 15% ó 20% de probabilidad!).

No debemos olvidar que los supuestos se refieren a las poblaciones y no a las muestras. Es bastante probable que una muestra no siga una curva normal aunque provenga de una población que sí lo hace. La figura 15-1 representa gráficamente histogramas de varias muestras, cada una tomada al azar de una población que sigue una distribución normal. (En la figura se observa que cuanto menor es la muestra, más difícil es notar que proviene de una población normal). Por supuesto, es bastante probable que poblaciones no normales produzcan cualquiera de esas muestras también. Lamentablemente, la muestra es todo lo que tenemos cuando realizamos un estudio. Lo que los investigadores hacen es trazar un histograma que represente la muestra y, si el gráfico no es significativamente diferente de lo normal, los investigadores suponen que la población de donde proviene es normal. Cuando se trata de normalidad, la mayoría de los investigadores psicológicos considera que una distribución es inocente hasta que se demuestre lo contrario.

Una situación común en la que un investigador podría dudar del supuesto en cuanto a que la población sigue una distribución normal, es cuando existe un efecto techo o piso (*véase* capítulo 1). Otra situación común que origina las mismas dudas es aquella en la que la muestra tiene valores atípicos, casos extremos a uno o a ambos lados de la distribución muestral. La figura 15-2 representa gráficamente algunos ejemplos de distribuciones con casos atípicos. Los casos atípicos son un gran problema para los métodos estadísticos que utilizamos comúnmente, ya que estos se basan, en última instancia, en desvíos cuadráticos de la media. Al encontrarse tan lejos de la media, un caso atípico tiene una gran influencia cuando elevamos al cuadrado su desvío con respecto a la media. El resultado es que un sólo caso atípico, si es lo suficientemente extremo, puede hacer que una prueba estadística resulte significativa aun cuando todos los otros valores no lo harían. Un caso atípico también puede hacer que un resultado que sin él sería significativo, no lo sea.

[1] Al calcular la significación de la regresión (capítulo 4) suponemos que en la población, para cada nivel de la variable de predicción, la variable dependiente es normal. También suponemos que la varianza de la variable dependiente es la misma para cada nivel de la variable de predicción. En la correlación (capítulo 3), el requisito es aún más estricto, siendo necesario que cada variable y las combinaciones de variables tengan distribuciones normales. Textos más avanzados presentan métodos sofisticados para identificar si se cumplen los supuestos mencionados. Sin embargo, al menos podemos considerar que los supuestos *no* han sido cumplidos *si* los datos de la muestra sugieren que en la población la distribución general con respecto a la variable dependiente (en la regresión), o a ambas (en la correlación), no es normal.

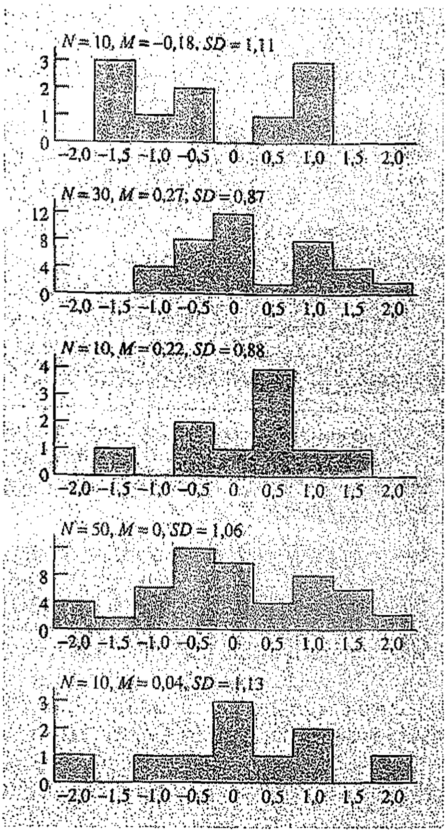


Figura 15-1.
Histogramas de varias muestras elegidas al azar, tomadas cada una de una población normal con $\mu = 0$ y $\sigma = 1$.

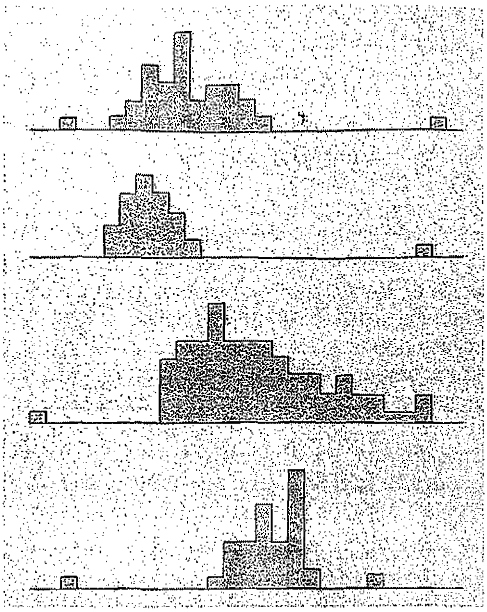


Figura 15-2.
Distribuciones con casos atípicos en uno o ambos lados.

TRANSFORMACIONES DE DATOS

Un procedimiento ampliamente utilizado cuando los valores de la muestra no parecen provenir de una población normal ¡es cambiar los valores! Por supuesto que no se inventan, aunque eso puede parecer antes de que expliquemos el procedimiento. El método consiste en que el investigador aplique algún procedimiento matemático a cada valor, como calcular la raíz cuadrada, para hacer que una distribución no normal se acerque más a lo normal. (Algunas veces este procedimiento también logra que las varianzas de dos o más grupos se asemejen más). El proceso que describimos en el párrafo anterior se denomina **transformación de datos**. Una vez que hemos realizado una transformación de datos, si se cumplen los otros supuestos podemos entonces calcular una prueba *t*, un análisis de varianza o una correlación ordinaria y, así, obtener resultados precisos.

La transformación de datos tiene una ventaja importante con respecto a otros procedimientos que aprenderemos para trabajar con poblaciones no normales: una vez que hemos realizado una transformación de datos, podemos utilizar procedimientos familiares y sofisticados de prueba de hipótesis.

Analicemos un ejemplo. Las medidas de tiempo de reacción usualmente son muy asimétricas hacia la derecha. Hay muchas respuestas cortas (rápidas) y unas pocas, pero a veces muy extremas, largas (lentas). Es improbable que los tiempos de reacción que aparecen en la figura 15-3 provengan de una población que sigue una curva normal; en realidad es probable que la propia población de tiempos de reacción sea asimétrica.

Sin embargo, supongamos que sacamos la raíz cuadrada de cada tiempo de reacción. La mayoría de los tiempos de reacción serán apenas afectados. Un tiempo de reacción de 1 segundo continúa siendo 1; un tiempo de reacción de 1,5 segundos se reduce a 1,22. Pero los tiempos de reacción muy lentos, los que crean la larga cola hacia la derecha, son reducidos sustancialmente. Por ejemplo, un tiempo de reacción de 9 segundos se reduce a 3, y un tiempo de reacción de 16 segundos (la persona realmente estaba distraída y se olvidó de la tarea que estaba realizando) se reduce a 4. La figura 15-4 muestra el resultado después de sacar la raíz cuadrada de cada tiempo de la distribución asimétrica representada en la figura 15-3. Después de una **transformación raíz cuadrada**, parece mucho más probable que la distribución de los tiempos de reacción provenga de una población con una distribución normal (de valores transformados).

Legitimidad de la transformación de datos

¿Es un engaño el proceso que describimos? Lo sería si se realizara sólo con algunos valores o de algún otro modo que hiciera el resultado más favorable a la predicción del investigador. Sin embargo, en las investigaciones reales, el primer paso después de recopilar y registrar los datos (y controlar su precisión) es observar si sugieren que las poblaciones cumplen con los supuestos. Si los datos sugieren que las poblaciones no cumplen con los supuestos, entonces el investigador realiza transformaciones de datos. La prueba de hipótesis se realiza sólo después del control mencionado y cualquier transformación necesaria.[2]

No debemos olvidar que cualquier transformación de valores debe realizarse con **todos** los valores de esa variable, no sólo con los de un subgrupo determinado. Lo más importante es que, cualquiera sea el procedimiento de transformación que utilicemos, el orden de los valores siempre es el mismo. Un valor original, que es el segundo más alto de un grupo de valores, continuará siendo el segundo más alto en el grupo de valores transformados.

Puede parecer que el procedimiento distorsiona de algún modo la realidad para adecuarla a la estadística. En algunos casos, esta es una preocupación legítima. Supongamos que se analiza la diferencia de ingresos entre dos grupos. Probablemente no nos interese cuánto difieren los

[2] Una vez que se completa un estudio, y antes de realizar cualquier cálculo estadístico descriptivo o prueba de significación, los investigadores primero controlan cuidadosamente que toda la información haya sido registrada correctamente e ingresada con precisión en la computadora. Después controlan cada variable en cuanto a la forma de su distribución, para ver si su población difiere seriamente de lo normal, proceso que se denomina "exploración de datos". La exploración de datos es un trabajo tedioso, y los investigadores están naturalmente ansiosos de encontrar lo antes posible la forma en la que funciona el estudio. Sin embargo, los investigadores experimentados han aprendido que vale la pena posponer la prueba de hipótesis para realizar un buen análisis exploratorio de datos. Es muy frustrante realizar todo tipo de análisis y luego descubrir que el trabajo ha sido una pérdida de tiempo porque había un error en el ingreso de la información o porque una de las variables necesitaba ser transformada. (De hecho, es peor que una pérdida de tiempo. El investigador puede entusiasmarse o desanimarse mucho con sus supuestos resultados y luego descubrir que las conclusiones no tenían sentido y que debe comenzar todo el proceso nuevamente).

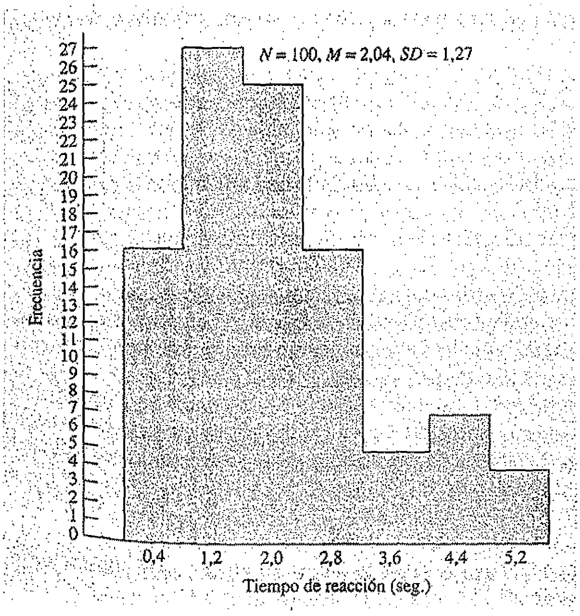


Figura 15-3.
Distribución asimétrica de tiempos de reacción (datos ficticios).

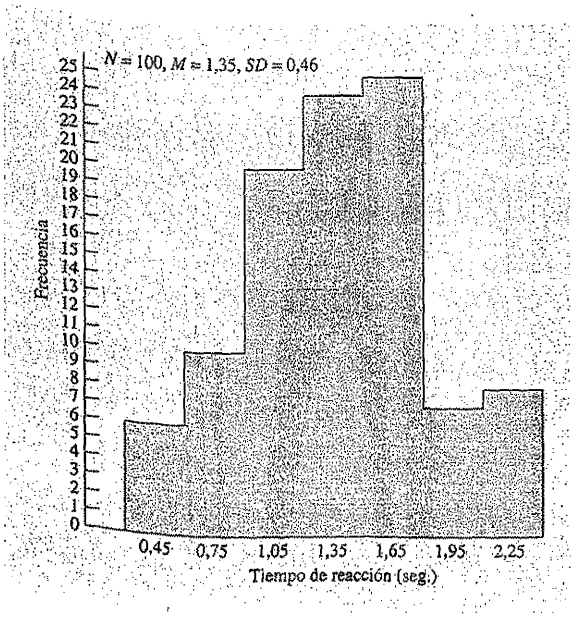


Figura 15-4.
Datos de la figura 15-3 después de la transformación raíz cuadrada.

dos grupos con respecto a la raíz cuadrada de sus ingresos, sino que lo importante es la diferencia en dólares reales.

Por otro lado, analicemos el cuestionario sobre autoestima. Las puntuaciones del cuestionario no tienen ningún significado directo. Las puntuaciones más altas indican mayor autoestima; las puntuaciones más bajas, menor autoestima. Sin embargo, cada unidad de aumento en la prueba no necesariamente está relacionada con una cantidad igual de aumento en la autoestima de un individuo. Es verosímil que la raíz cuadrada de cada incremento de una unidad esté directamente relacionada con la autoestima de una persona. De modo similar, si analizamos el ejemplo utilizado anteriormente acerca del tiempo de reacción, medido en segundos, la medición parecería tener un significado directo. Sin embargo, incluso en ese caso, la variable implícita, eficiencia del procesamiento del sistema nervioso, puede no estar directamente relacionada con la cantidad de segundos. Probablemente es una operación compleja que sigue alguna regla matemática desconocida (aunque siempre esperaríamos que los tiempos más cortos indicaran un procesamiento más eficiente, y tiempos más largos un procesamiento menos eficiente).

En los ejemplos anteriores, el "patrón" implícito de la variable es desconocido. Por lo tanto, no existe razón para pensar que la versión transformada sea un reflejo menos preciso de la realidad que la versión original. Y la versión transformada puede cumplir el supuesto de normalidad.

Tipos de transformaciones de datos

Existen distintos tipos de transformaciones de datos. Ya hemos ilustrado una transformación raíz cuadrada: en lugar de utilizar cada valor, se utiliza su raíz cuadrada. Las figuras 15-3 y 15-4 ilustran un ejemplo, y la figura 15-5 representa gráficamente el efecto general: una distribución asimétrica hacia la derecha se transforma en una distribución menos asimétrica hacia la derecha después de la transformación raíz cuadrada. Para expresarlo numéricamente, los números moderados se convierten en números levemente menores y los números altos se convierten en números mucho menores. El resultado es que el lado derecho es empujado hacia el centro.

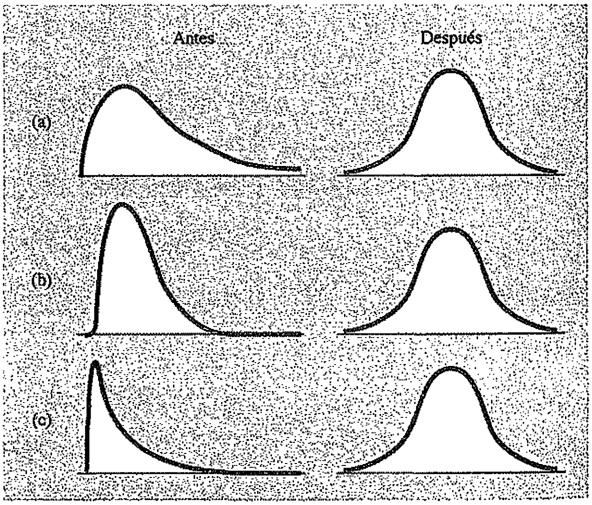


Figura 15-5.
Distribuciones a las que se les aplican las transformaciones apropiadas: (a) moderadamente asimétrica hacia la derecha, se aplica a la transformación raíz cuadrada; (b) marcadamente asimétrica hacia la derecha, se aplica a la transformación *log*, y (c) extremadamente asimétrica hacia la derecha, se aplica a la transformación inversa.

¿Qué sucede si la distribución es asimétrica hacia el otro lado (hacia la izquierda)? En este caso, primero podemos **reflejar** todos los valores, es decir, restarlos a un número alto de modo que todos se reviertan. Después de reflejar los valores, una distribución que era asimétrica hacia la izquierda se transforma en asimétrica hacia la derecha, y una transformación raíz cuadrada producirá el efecto correcto. Sin embargo, cuando reflejamos los valores, al analizar los resultados finales debemos recordar que hemos revertido la dirección de los valores. Lo que se solía considerar un valor alto ahora es un valor bajo, y viceversa.

Otra transformación muy común es la **transformación *log***. Una transformación *log* tiene el mismo efecto general que la transformación raíz cuadrada. Hace que una distribución asimétrica hacia la derecha sea menos asimétrica hacia la derecha. Pero la transformación *log* es más severa. Puede convertir en normal a una distribución incluso más extremadamente asimétrica. La figura 15-5b representa gráficamente la situación descripta.

El alumno seguramente recordará, de las clases de matemáticas de la escuela secundaria, que un logaritmo es el exponente al que se debe elevar un número base (como por ejemplo 10) para obtener el número original. Por ejemplo, el logaritmo d 100 de base 10 es 2; para obtener el número 100 debemos elevar 10 a la segunda potencia (lo elevamos al cuadrado). En otras palabras, 2 es el valor correspondiente a 100 después de una transformación *log* (utilizando logaritmos con base 10). El logaritmo de 1.000 sería 3; 10 a la tercera potencia (al cubo) es 1.000; 3 es el valor correspondiente a 1.000 después de una transformación *log*. El logaritmo de 10 es 1 (cualquier número a la primera potencia es el mismo número) y el logaritmo de 1 es 0 (cualquier número elevado a 0 es 1). El valor 10 se transforma en 1 y el valor 1 se transforma en 0. Los números intermedios tienen logaritmos con cifras decimales. El logaritmo de 50 es 1,70; de 60 es 1,78; de 8 es 0,90, y de 328 es 2,52. Un rango de 1 a 1.000 se ha reducido a un rango de 0 a 3, siendo el efecto mucho mayor cuanto más altos son los números.

No es necesario calcular logaritmos, el cálculo se realiza con cualquier calculadora. Uno de los aspectos más importantes que debemos recordar es que una transformación *log* produce exactamente el mismo efecto que una transformación raíz cuadrada, sólo que en mayor grado. Se aplicaría cuando la distribución de los datos fuera tan asimétrica hacia la derecha que una transformación a la raíz cuadrada tampoco puede convertir la distribución en una distribución aproximadamente normal.

Otro tipo común de transformación es la **transformación inversa**. En este caso, se toma el número inverso al valor, es decir, se lo convierte en el denominador de una fracción en la que el numerador es 1. El inverso de 10 es 1/10 (0,1); el inverso de 5 es 1/5 (0,2); el inverso de 1.000 es 1/1.000 (0,001). Lo importante es que una transformación inversa también produce el mismo efecto que las transformaciones raíz cuadrada y *log*, pero es aún más extrema que la transformación *log*. La transformación inversa es útil para datos demasiado asimétricos, incluso para ser convertidos en normales por una transformación *log*. La figura 15-5c representa gráficamente esta situación.

Además, la transformación inversa automáticamente revierte la dirección de los registros. Por ejemplo, en términos de valores originales, 5 es menor que 10. Después de una transformación inversa, el orden se revierte. La versión invertida de 5 es 1/5 (0,2), que es un número mayor a la versión invertida de 10, que es 1/10 (ó 0,1). Para mantener los datos en orden, los investigadores algunas veces reflejan los valores antes o después de una transformación inversa. Como una doble negación, el proceso reubica los datos en su dirección original.

Existen otras transformaciones. Por un lado, todas las transformaciones que hemos analizado hasta ahora corrigen una distribución asimétrica. Otras transformaciones tratan problemas de curtosis y de distribuciones "abultadas". Las distribuciones basadas en proporciones o porcentajes a menudo están lejos de ser normales, pero pueden corregirse con lo que se denomina una distribución arco-seno. Se trata de una función trigonométrica disponible en algunas calculadoras y en la mayoría de los programas de estadística para computadoras. Otras transformaciones que podemos encontrar son las transformaciones "*logit*" y "*probit*", al igual que transformaciones a distintas potencias, tales como transformaciones cuadradas o cúbicas.

No daremos ejemplos de todos estos otros tipos de transformaciones. Aprender las transformaciones cuadrada, *log* e inversa ayudará a captar el principio básico y, además, son las transformaciones más comunes. Lo principal acerca de los otros tipos de transformaciones es que todas utilizan el mismo principio en cuanto a tomar cada valor y aplicarle algún cálculo aritmético, usualmente para que la serie de valores sea más parecida a una distribución normal. Una vez más, cualquiera sea la transformación utilizada, un valor que se encuentra entre otros dos valores siempre permanece entre esos mismos dos valores.

Ejemplo de transformación de datos

Analicemos un estudio ficticio en el que cuatro niños con valores altos en una prueba sobre "alta sensibilidad" son comparados, en cuanto a la cantidad de libros leídos durante el año anterior, con cuatro niños que tuvieron valores bajos en la misma prueba. (La noción general de persona altamente sensible se describe en Aron, 1996 y Aron & Aron, 1997). Basándonos en la teoría, el investigador predice que los niños altamente sensibles habrán leído más libros. La tabla 15-1 refleja los resultados.

Comúnmente, en un estudio de este tipo, que incluye una comparación de dos grupos independientes, deberíamos utilizar una prueba *t* para medias independientes. Pero la prueba *t* para medias independientes es igual a todos los procedimientos de prueba de hipótesis aprendidos (excepto el chi-cuadrado): requiere que la población madre de observaciones de cada grupo esté normalmente distribuida. En este estudio, sin embargo, la distribución de la muestra es muy asimétrica hacia la derecha, y los valores tienden a acumularse a la izquierda formándose una larga cola hacia la derecha. Por eso parece probable que la población de observaciones de cantidad de libros leídos (tanto para niños sensibles como para los no sensibles) también sea asimétrica hacia la derecha. Además, a la luz de lo que se está midiendo, la forma descripta de la distribución poblacional parece razonable: un niño no puede leer menos que ningún libro; pero una vez que un niño comienza a leer, es fácil que lea muchos libros en un año.

Tabla 15-1.
Resultados de un estudio que compara niños altamente y no altamente sensibles con relación a la cantidad de libros leídos durante el año anterior (datos ficticios).

	Altamente sensible	
	No	*Si*
	0	17
	3	36
	10	45
	22	75
Σ:	35	173
$M =$	8,75	43,25
$S^2 =$	95,58	584,00

También podemos observar que las varianzas poblacionales estimadas sobre la base de las dos muestras son significativamente diferentes, 95,58 contra 584, otra razón para no querer proseguir con una prueba *t* ordinaria.

Sin embargo, supongamos que realizamos una transformación raíz cuadrada de las observaciones (tabla 15-2). El resultado es que ambas muestras son mucho más adaptables a una curva normal, y la transformación también parece razonable en cuanto al significado de los números. La cantidad de libros leídos pretende ser una medida del interés literario; por lo tanto, la diferencia entre 0 y 1 libro es una diferencia mucho mayor que la que existe entre 20 y 21 libros.

La tabla 15-3 muestra la prueba *t* utilizando los valores transformados. Como lo indica la tabla, la diferencia entre los grupos es significativa.[3]

Otro ejemplo de transformación de datos

Esta vez analizaremos un ejemplo ficticio que incluye una correlación. Si el alumno aún no ha estudiado el capítulo 3, debería pasar por alto esta sección.

Un psicólogo especializado en educación realiza un estudio acerca de la relación entre la puntuación en una prueba de álgebra y la nota de nivel escolar, probando a cuatro alumnos de la escuela. La figura 15-6 muestra el diagrama de dispersión; la tabla 15-4 muestra las observaciones y los cálculos del coeficiente de correlación.

Como se desprende de la figura y de la tabla, existe cierto grado de correlación. Sin embargo, si observamos nuevamente la lista de valores observados y el diagrama de dispersión, el niño con 95 puntos en la prueba de álgebra sobresale entre los demás. El valor 95 es un caso atípico. Para decirlo de otro modo, las puntuaciones de la prueba de álgebra son asimétricos hacia la derecha; se agrupan cerca del límite inferior de las puntuaciones de la prueba y tienen una larga cola que llega al valor 95, que se ubica lejos a la derecha. Las notas de nivel escolar no son muy diferentes de lo que se hubiera esperado si provinieran de una población normal. Las notas se agrupan cerca de la mitad (el 6 y el 7) y luego se dispersan un poco más, en forma pareja, para ambos lados (el 4 y el 9).

Tabla 15-2.
Transformación raíz cuadrada de los registros de la tabla 15-1.

Altamente sensibles			
No		*Sí*	
X	$\sqrt{X}$	X	$\sqrt{X}$
0	0,00	17	4,12
3	1,73	36	6,00
10	3,16	45	6,71
22	4,69	75	8,66

[3] Si hubiéramos realizado el análisis utilizando los valores originales sin transformar, *t* sería igual a (43,25 – 8,75)/13,04 ó 2,65, un *t* levemente menor pero aún significativo. Por supuesto, no hubiera sido correcto realizar el análisis de ese modo. Si el análisis realizado con valores no transformados hubiera producido un resultado diferente, el resultado correcto hubiera sido el logrado sobre la base de los valores transformados.

Tabla 15-3.
Cálculos de una prueba t para medias independientes aplicando la transformación raíz cuadrada a los valores observados del estudio acerca de los libros leídos por niños altamente sensibles contra los no altamente sensibles (datos ficticios).

Punto de corte t para nivel 0,05 de significación, $gl = (4 - 1) + (4 - 1) = 6$, una cola $= -1,943$.

	Altamente sensible	
	No	*Sí*
	0,00	4,12
	1,73	6,00
	3,16	6,71
	4,69	8,66
Σ:	9,58	25,49
$M =$	9,58/4 = 2,40	25,49/4 = 6,37
$S^2 =$	12,03/3 = 4,01	10,56/3 = 3,52

$S^2_{\text{combinado}} = 3,77$

$S^2_M =$	3,77/4 = 0,94	3,77/4 = 0,94

$S^2_{\text{diferencia}} = 0,94 + 0,94 = 1,88$
$S_{\text{diferenica}} = \sqrt{1,88} = 1,37$
$t = (6,37 - 2,40)/1,37 = 2,90$

Conclusión: se rechaza la hipótesis nula.

¿Qué debemos hacer entonces con la distribución de la prueba de álgebra? Primero, en casos como este, en los que hay un sólo y claro caso atípico, probablemente deberíamos controlar que no existan errores de calificación o intentar averiguar si ese individuo en particular era de algún modo atípico con respecto a la población bajo estudio (como por ejemplo, alguien que estuviera en un programa acelerado de aprendizaje de matemática o cuya madre fuera una matemática famosa). Sin embargo, suponiendo que nada se sabe, ni se puede averiguar, la otra solución es transformar las puntuaciones de la prueba para que no sean asimétricas. Además, la segunda alternativa también resulta razonable en este caso, ya que no tiene ningún valor especial conocer la cantidad original de los ítems correctamente respondidos en la prueba.

Los valores son asimétricos hacia la derecha, así que probablemente necesitemos utilizar una transformación raíz cuadrada, *log* o inversa. Podemos comenzar intentando una transformación raíz cuadrada. A través de la transformación mencionada, las puntuaciones de la prueba de álgebra se transforman de 1, 4, 10 y 95 a 1, 2, 3,2 y 9,7. La situación ha mejorado pero continúa siendo bastante asimétrica hacia la derecha. Se necesita una transformación más extrema. Podríamos intentar una transformación *log*. Utilizando una calculadora (con tecla para logaritmo con base 10), calculamos los logaritmos para 1, 4, 10 y 95. Los resultados fueron 0, 0,6, 1 y 1,98. Esta vez la distribución resultante es sólo levemente asimétrica hacia la derecha, y parece ser una probable candidata a la muestra seleccionada de una población (de valores transformados a logaritmos), en la que la mayoría de los valores se agrupan en el medio y hay una cantidad menor, pero pareja, de valores en los dos extremos.

Habiendo encontrado una transformación adecuada, ahora podemos intentar nuevamente nuestra correlación. La figura 15-7 muestra el diagrama de dispersión, y la tabla 15-5 muestra los cálcu-

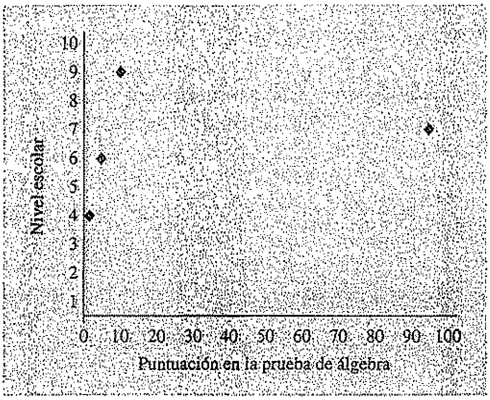


Figura 15-6.
Diagrama de dispersión de un estudio acerca de la nota de nivel escolar y la puntuación en una prueba de álgebra. (Datos ficticios).

Tabla 15-4.
Registros y cálculos de un estudio que correlaciona las notas de nivel escolar y las puntuaciones en una prueba de álgebra (datos ficticios).

	Puntuación en la prueba		Nota de nivel escolar		Producto cruzado
	Original	Z_X	*Original*	Z_Y	$Z_X Z_Y$
	1	−0,68	4	−10,47	1,00
	4	−0,60	6	−0,29	0,17
	10	−0,45	9	1,47	−0,66
	95	10,73	7	0,29	0,50
Σ:	110		26		1,01
$M =$	27,5		6,5		$r =$ 0,25
$SD =$	39,1		1,7		

los. La correlación calculada con los valores transformados a logaritmos es de 0,65. La correlación calculada utilizando valores no transformados, como lo muestra la tabla 15-4, es de sólo 0,25.

La tabla 15-5 también indica los cálculos de la significación del coeficiente de correlación (el procedimiento está tomado del apéndice II del capítulo 3). Incluso con una correlación tan alta, como de 0,65, con sólo cuatro participantes existe insuficiente potencia para rechazar la hipótesis nula. (Cuando se trabaja con una correlación, la hipótesis nula establece que la correlación en la población es 0). Sin embargo, al menos fue correcto calcular esta prueba t, en el sentido de que habíamos cumplido el supuesto de distribuciones normales. (Supongamos que hubiéramos calculado incorrectamente t para la correlación de 0,25 a partir de los valores sin transformar. El t hubiera sido de sólo 0,37, contra 1,21 con los valores transformados).

PRUEBAS DE RANGO Y ORDEN

Otra forma de solucionar el problema de las distribuciones no normales es utilizar un tipo especial de transformación en la que los valores se transforman en rangos. Supongamos que tenemos una muestra con los valores 4, 8, 12 y 64. Sería en verdad sorprendente obtener una muestra así si la población fuera realmente normal. Una **transformación de rango y orden** cambiaría los valo-

Figura 15-7.
Diagrama de dispersión del estudio acerca de las notas de nivel escolar y las puntuaciones en una prueba de álgebra representado por la figura 15-6, después de una transformación *log* de las puntuaciones en la prueba de álgebra (datos ficticios).

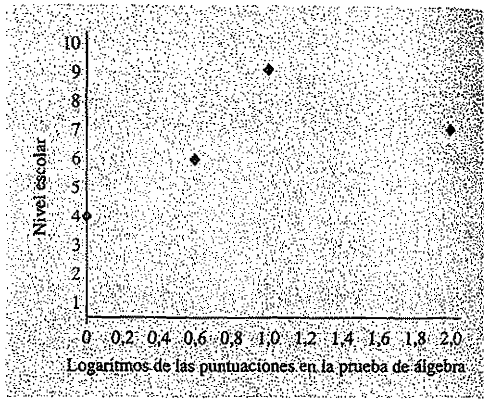


Tabla 15-5.
Observaciones y cálculos de un estudio que correlaciona la nota de nivel escolar y las puntuaciones en una prueba de álgebra transformadas a logaritmos (datos ficticios).

	Puntuación en la prueba		Nota de nivel escolar		Producto cruzado
	Original	Z_X	*Original*	Z_Y	Z_XZ_Y
	0,00	−1,25	4	−1,47	1,84
	0,60	−0,42	6	−0,29	0,12
	1,00	0,14	9	1,47	0,21
	1,98	1,50	7	0,29	0,44
Σ:	3,58		26		2,61
M =	0,90		6,5		*r* = 0,65
SD =	0,72		1,7		

Prueba de significación:

Punto de corte *t* necesario (nivel 0,05, $gl = 2$, una cola) = 2,920
$t = (r\sqrt{N-2})/\sqrt{1-r^2} = (0{,}65)(1{,}41)/\sqrt{1-0{,}65^2} = 0{,}92/0{,}76 = 1{,}21$

Conclusión: no se rechaza la hipótesis nula que establece que $r = 0$.

res a 1, 2, 3 y 4; el 1 para el número más bajo del grupo, el 2 para el siguiente más bajo, y así sucesivamente.

La única complicación de la transformación de rango y orden surge cuando existen dos o más valores iguales. La solución usual para los casos en los que existen valores iguales es darle a cada uno el promedio de los rangos correspondientes. Por ejemplo, a los valores 12, 81, 81, 107 y 154 les corresponderían los rangos 1, 2,5, 2,5, 4 y 5.

Convertir los valores en rangos es una especie de transformación de datos, pero a diferencia de las transformaciones que hemos analizado hasta ahora, una transformación de rango y orden no se utiliza para producir una distribución normal, aunque, en efecto, produce una distribución particular. La distribución que se obtiene a partir de una transformación de rango y orden es rectangular, con la misma cantidad de valores (uno) para cada valor (la única excepción son los valores iguales). Los rangos producen el efecto de dispersar los valores en forma pareja.

Existen diversos procedimientos especiales de prueba de hipótesis que utilizan datos transformados en rangos. Se los denomina **pruebas de rango y orden.** También tienen otros dos nombres comunes: dado que los datos de una población con cualquier tipo de distribución pueden transformarse en rangos, estas pruebas a veces se denominan **pruebas libres de** distribución; y dado que la distribución de valores convertidos en rangos no es estimada sinó que se conoce con exactitud, las pruebas de rango y orden no requieren la estimación de ningún parámetro (valores de la población). (Por ejemplo, no hace falta estimar la varianza de una población porque podemos determinarla exactamente si sabemos cuántos valores la forman y que esos valores han sido transformados en rangos). Por eso, los procedimientos de prueba de hipótesis basados en rangos también se denominan **pruebas no paramétricas.**

Los procedimientos ordinarios de prueba de hipótesis que hemos aprendido (prueba *t* y análisis de varianza) son ejemplos de **pruebas paramétricas.** El chi-cuadrado, al igual que las pruebas de rango y orden, se considera una prueba no paramétrica; sin embargo, es libre de distribución sólo en el sentido de que no existen supuestos sobre la forma de las distribuciones poblacionales. No obstante, los términos **libre de distribución y no paramétrico** generalmente se utilizan en forma indistinta; las sutilezas con respecto a la diferencia entre esos términos son materia de debate entre los estadísticos.

Las pruebas de rango y orden tienen la ventaja adicional de poder utilizarse cuando los valores reales del estudio son rangos; por ejemplo, un estudio que compara el nivel social de dos clases de graduados. Además, algunas veces son cuestionables los valores numéricos exactos de los números de una medida utilizada en determinado estudio. Por ejemplo, un investigador tiene la intención de aplicar una medida numérica en el sentido usual, siendo 7 tan superior a 5 como 12 lo es de 10 (el investigador pretende que ésta sea una "medición intervalar"; *véase* capítulo 1). Sin embargo, en realidad sólo está seguro de que los números están ordenados correctamente: 7 es mayor que 5, 10 es mayor que 7, y así sucesivamente. En ese caso, el investigador podría utilizar una medición de rango y orden para no sobrestimar la calidad del instrumento o procedimiento de medición.

En realidad, el tema es algo controvertido. Analicemos, por ejemplo, una escala en la que 1 = en desacuerdo; 2 = medianamente en desacuerdo; 3 = medianamente de acuerdo, y 4 = de acuerdo. Los significados implícitos en los números, ¿están dispersos en forma pareja en la escala numérica? Queda claro que los resultados tienen sentido como datos de rango y orden –ciertamente, 2 muestra más aprobación que 1, 3 más que 2 y 4 más que 3. Por eso, algunos psicólogos sostienen que, en la mayoría de los casos, no deberíamos suponer que tenemos mediciones intervalares, y deberíamos convertir nuestros datos en rangos y utilizar una prueba de significación de rango y orden. Otros investigadores sostienen que las pruebas estadísticas paramétricas resultan razonablemente precisas incluso con mediciones de rango y orden reales, y que al cambiar todos los datos a rangos se puede perder información valiosa. La cuestión sigue sin resolverse.

Idea general acerca de las pruebas de rango y orden

La tabla 15-6 muestra el nombre de las pruebas de rango y orden con las que se sustituiría cada procedimiento paramétrico ordinario de prueba de hipótesis que hemos aprendido. Cuando se indica más de una prueba posible, los procedimientos son aproximadamente equivalentes.[4]

A continuación describimos en forma general el modo en que se realizan estas pruebas, incluyendo un ejemplo. Sin embargo, no daremos toda la información necesaria para realizar una de esas pruebas en la práctica; sólo presentamos estas técnicas porque seguramente aparecerán en

[4] Existe una prueba no paramétrica ampliamente utilizada, además de las pruebas chi-cuadrado, que no se basa en registros de rango y orden. Se la denomina **prueba de signos.** Una prueba de signos se utiliza en lugar de una prueba *t* para medias dependientes. Se crea la serie de valores diferenciales y luego se suman sólo los números positivos.

publicaciones científicas, y porque su lógica es la base de un procedimiento alternativo que sí enseñaremos a utilizar. Ese procedimiento alternativo tiene casi la misma función que las pruebas de rango y orden, y es más parecido a las técnicas ya aprendidas.

Lógica básica de las pruebas de rango y orden

Analicemos un estudio que incluye un grupo experimental y un grupo control. (Es una situación típica en la que, si se cumplieran todos los supuestos, los psicólogos utilizarían una prueba *t* para medias independientes). Si quisiéramos utilizar una prueba de rango y orden, primero transformaríamos todos los valores observados en rangos, ordenando los valores de menor a mayor, sin importar si la observación pertenece al grupo experimental o de control. Si los dos grupos fueran de valores tomados al azar de una sola población, deberían haber aproximadamente las mismas cantidades de rangos altos y bajos en cada grupo (es decir, si la hipótesis nula es verdadera, los rangos de los dos grupos no deberían ser muy diferentes). Debido a que la distribución de rangos puede calcularse con exactitud, los estadísticos pueden calcular la probabilidad exacta de obtener cualquier división determinada de rangos en dos grupos si, de hecho, los dos grupos fueran tomados al azar de poblaciones idénticas.

En realidad, la forma para que esto funcione consiste en que el investigador convierta todos los valores observados en rangos, sume el total de los rangos del grupo con los valores menores y luego compare ese total con un punto de corte indicado en una tabla especial de estos puntos de corte de significación para totales de rangos en este tipo de situaciones.

Ejemplo de prueba de rango y orden

La tabla 15-7 muestra el cálculo de una prueba de suma de rangos de Wilcoxon para el tipo de situación que describimos anteriormente. Este ejemplo utiliza la misma información que nuestro primer ejemplo de transformación de datos, el estudio ficticio sobre cantidad de libros leídos por niños altamente sensibles comparado con los leídos por niños no altamente sensibles. La lógica es algo diferente, por eso recomendamos tener paciencia para esperar la explicación.

Tabla 15-6.
Principales pruebas de rango y orden equivalentes a las principales pruebas paramétricas.

Pruebas paramétricas ordinarias	Pruebas de rango y orden equivalentes
Prueba *t* para medias dependientes	Prueba de rango con signos de Wilcoxon
Prueba *t* para medias independientes	Prueba de suma de rangos de Wilcoxon o prueba *U* de Mann-Whitney
Análisis de varianza	Prueba *H* de Kruskal-Wallis
Prueba *t* para correlación	rho de Spearman o tau de Kendall

Si no existe diferencia promedio, aproximadamente la mitad de los valores diferenciales debería ser positiva y la mitad negativa. Si la cantidad de positivos es considerablemente mayor o considerablemente menor a la mitad, el resultado estaría en contra de una hipótesis nula que establece que la verdadera población de valores diferenciales tiene una diferencia promedio igual a cero. Los textos estadísticos de nivel intermedio usualmente incluyen una tabla donde buscar los puntos de corte de significación de una prueba de signos.

Como se desprende de la tabla, en primer lugar determinamos el punto de corte de significación, como haríamos en cualquier procedimiento de prueba de hipótesis (el punto de corte se basa en una tabla que no hemos proporcionado pero que se puede encontrar en la mayoría de los textos de estadística de nivel intermedio). El siguiente paso fue ordenar los rangos de menor a mayor; después, sumar el grupo que se espera que tenga el total más bajo. Luego, el total se compara con el punto de corte. En el ejemplo que analizamos, el total de los rangos del menor no fue mayor que el punto de corte; por lo tanto, se rechaza la hipótesis nula.

Utilizamos la prueba de suma de rangos de Wilcoxon, aunque podríamos haber utilizado la prueba *U* de Mann-Whitney, que da un resultado final matemáticamente equivalente y se basa en la misma lógica. Sólo difiere en los detalles de cálculo.

La hipótesis nula en una prueba de rango y orden

La hipótesis nula en una prueba de rango y orden no es exactamente igual a la de una prueba paramétrica ordinaria. Una prueba paramétrica compara las medias de los dos grupos; su hipótesis nula establece que las dos poblaciones tienen la misma media. En una prueba de rango y orden lo equivalente a la media es el rango medio (la mediana de los valores no convertidos a rango). Por ejemplo, supongamos que cinco valores no convertidos a rango fueran 11, 12, 14, 19 y 20. Sus rangos correspondientes son 1, 2, 3, 4 y 5. El rango medio es 3, que corresponde a la mediana de los valores no transformados a rangos, es decir, el valor 14. Por lo tanto, consideramos a una prueba de rango y orden como la comparación de medianas de los dos grupos, cuya hipótesis nula establece que las dos poblaciones tienen la misma mediana.

Aproximaciones a la curva normal en pruebas de rango y orden

Las tablas como las descriptas, para la suma máxima de rangos para rechazar la hipótesis nula, son muy engorrosas si se utilizan tamaños de muestras entre moderados y grandes, con grupos

Tabla 15-7.
Cálculos de una prueba de suma de rangos de Wilcoxon basados en el estudio acerca de los libros leídos por niños altamente sensibles en comparación con los leídos por niños no altamente sensibles (datos ficticios).

Punto de corte: suma máxima de rangos en el grupo no altamente sensible para un nivel 0,05 de significación, una cola (de una tabla estándar) = 11.

Altamente sensible			
No		*Sí*	
X	*Rango*	*X*	*Rango*
0	1	17	4
3	2	36	6
10	3	45	7
22	5	75	8
Σ:	11		

Comparación con el punto de corte: la suma de rangos del grupo que se predijo tendría los registros más bajos; 11, iguala pero no excede al punto de corte de significación.

Conclusión: se rechaza la hipótesis nula.

desiguales. Y el problema se torna realmente inmanejable con diseños más complicados. Por eso se han desarrollado varias aproximaciones que utilizan las sumas de rangos en una fórmula que produce una puntuación Z. Si la puntuación Z se encuentra en la región superior sobre la cual está el 5% del área, bajo la curva normal (2,5% para una prueba de dos colas), el resultado se considera significativo. Con frecuencia, cuando las publicaciones científicas informan las pruebas de rango y orden indican la puntuación *Z* que mencionamos.

Utilización de pruebas paramétricas con datos transformados en rangos

Conover e Iman (1981) demostraron que no es necesario realizar los procedimientos de cálculo especiales de las pruebas de rango y orden. Se pueden obtener aproximadamente los mismos resultados si transformamos los valores observados en rangos y luego aplicamos la aritmética usual para calcular una prueba paramétrica ordinaria, como por ejemplo una prueba *t*. (El procedimiento descripto funciona en el caso de una prueba *t*, un análisis de varianza de un criterio, y en la significación del coeficiente de correlación. No funciona tan bien con el análisis de varianza de dos criterios).

El resultado de utilizar una prueba paramétrica con valores transformados en rangos no será tan preciso como los de la prueba paramétrica ordinaria o la prueba de orden y rango. No serán tan precisos como los de la prueba paramétrica ordinaria porque no se cumple el supuesto de distribuciones normales, ya que, en realidad, cuando se trabaja con rangos la distribución es rectangular. Tampoco serán tan precisos como los de las pruebas de orden y rango porque la prueba paramétrica utiliza la distribución *t* ó *F* en lugar de las tablas especiales que utilizan las pruebas de orden y rango, las cuales se basan en probabilidades exactas de obtención de ciertas divisiones de rangos. Sin embargo, la aproximación parece ser bastante buena.[5]

Ejemplo de prueba paramétrica ordinaria después de una transformación de rango y orden

La tabla 15-8 muestra los cálculos de una prueba *t* ordinaria para medias independientes realizada con los datos ficticios acerca de niños sensibles, utilizando el rango de cada niño en lugar del número real de libros leídos por ellos. Nuevamente, obtenemos un resultado significativo.

MÉTODOS INTENSIVOS POR COMPUTADORA

En los últimos años, gracias a la disponibilidad de computadoras ha adquirido practicidad toda una nueva serie de métodos de prueba de hipótesis. Las principales técnicas se denominan **pruebas de aleatorización** y "***boots trap***". Los métodos mencionados difieren en algunos detalles importantes. Sin embargo, su lógica es lo suficientemente parecida como para que podamos transmitir la idea básica concentrándonos en uno de ellos: las pruebas de aleatorización.

[5] Un investigador particularmente preocupado por la precisión podría calcular *t* ó *F* utilizando los valores transformados a rangos, y después convertir el resultado en el resultado exacto de una prueba de rango y orden, utilizando una fórmula de conversión establecida por Conover e Iman (1981). Luego buscaría ese número en la tabla apropiada de pruebas de rango y orden.

Tabla 15-8.
Cálculos de una prueba t para medias independientes utilizando rangos en lugar de los valores originales del estudio acerca de libros leídos por niños altamente sensibles en comparación con los leídos por niños no altamente sensibles (datos ficticios).

Punto de corte t para el nivel 0,05 de significación, $gl = (4 - 1) + (4 - 1) = 6$, una cola $= -1,943$

	Altamente sensible	
	No	Sí
	1	4
	2	6
	3	7
	5	8
Σ	11	25
$M =$	$11/4 = 2,75$	$25/4 = 6,25$
$S^2 =$	$8,75/3 = 2,92$	$8,75/3 = 2,92$

$$S^2_{\text{Combinada}} = 2,92$$

$S^2_M =$	$2,92/4 = 0,73$	$2,92/4 = 0,73$

$S^2_{\text{Diferencia}} = 0,73 + 0,73 = 1,46$

$S_{\text{Diferencia}} = \sqrt{1,46} = 1,21$

$t = (2,75 - 6,25)/1,21 = -2,89$

Conclusión: se rechaza la hipótesis nula.

Prueba de aleatorización básica

Supongamos que tenemos dos grupos de observaciones, uno que proviene de un grupo experimental y otro de un grupo de control. Supongamos también que las medias de los dos grupos difieren en cierto grado. Ahora analicemos qué sucede si mezcláramos todas esas observaciones ignorando de qué grupo provienen. Si calculáramos la diferencia entre las medias de esos dos grupos establecidos al azar, ¿cuál es la probabilidad de que todo ese proceso diera como resultado una diferencia media tan grande como la encontrada originalmente por la propia agrupación de las observaciones? Si la diferencia media entre los grupos originales es bastante pequeña, es muy probable que podamos obtener una diferencia media de ese tamaño a través de las agrupaciones casuales. Pero si la diferencia media de los grupos originales es bastante grande, al crear grupos al azar por lo general no obtendremos una diferencia de igual tamaño. Si las agrupaciones por azar produjeran un resultado del tamaño de las agrupaciones originales menos del 5% de las veces, podríamos estar bastante seguros de que las agrupaciones originales eran bastante diferentes de lo que esperaríamos por casualidad. Por lo tanto, el hecho de comparar agrupaciones reales de observaciones con agrupaciones aleatorias de observaciones es una forma de realizar una prueba de significación.

Una prueba de aleatorización, en este tipo de situaciones, en realidad establece rápidamente, por computadora, cada una de las posibles divisiones de las observaciones en dos grupos de los tamaños pertinentes. Luego determina cuántas de esas posibles organizaciones presentan una diferencia tan extrema como las diferencias realmente observadas entre los dos grupos. Si menos del 5% de las posibles organizaciones arrojan diferencias tan extremas como las originales, el resultado es significativo. Se puede rechazar la hipótesis nula que establece que los dos grupos podrían haber presentado esa diferencia en una división aleatoria. (La lógica descripta es similar a la utilizada para descubrir las probabilidades en las pruebas de rango y orden, pero, en este caso, las observaciones no fueron convertidas primero en rangos).

Ejemplo de prueba de aleatorización

La tabla 15-9 muestra un ejemplo resuelto de lo que realmente haría una computadora en el caso de una prueba de aleatorización aplicada al ejemplo del estudio ficticio acerca de dos grupos y de la cantidad de libros leídos.

A continuación describimos los pasos de una prueba de aleatorización con respecto a la diferencia entre las medias de dos grupos. (No debemos olvidar que en los casos reales las computadoras realizan todo el proceso. Sin embargo, tanto aquí como en los ejercicios seguimos el proceso paso a paso, para que el alumno pueda comprender rápidamente los resultados del procedimiento al verlos en las impresiones emitidas por las computadoras o en las publicaciones científicas. En el futuro, este puede convertirse en un método de uso muy común en la medida en que los investigadores psicológicos se adapten a todas las posibilidades ofrecidas por las computadoras de alta velocidad.

1. Determinar la diferencia entre las medias de los dos grupos reales. En el ejemplo, la diferencia media de cantidad de libros leídos era 34,5.

2. Determinar cuántas maneras posibles de formar dos grupos hay y cuán alta debería ser la diferencia media real para estar en el 5% (ó 1%) superior. Existen reglas que determinan cuántas posibles combinaciones serían necesarias. Esas reglas se describen en algunos textos de estadística de nivel intermedio, tratadas como permutaciones y combinaciones. Ya que en investigaciones reales nunca realizaríamos manualmente una prueba de aleatorización, dejaremos el tema de las reglas para futuros cursos. (En los ejercicios de este capítulo pedimos al alumno que calcule algunas pruebas de aleatorización con pequeñas cantidades de valores, con el fin de que incorpore el principio de las mismas. En esos casos, indicaremos cuántas combinaciones son necesarias).

En el caso que venimos analizando, con 8 participantes divididos en dos grupos de cuatro existen 70 maneras posibles de formar los grupos. Cada una de esas 70 divisiones produce una diferencia entre las medias de los dos grupos resultantes formados por cuatro registros. De esas 70 diferencias, el 5% superior son las 3,5 diferencias superiores. La diferencia media real tendrá que estar entre las tres superiores para que podamos rechazar la hipótesis nula (a menos que en el tercer lugar coincidan más de una media). (Si quisiéramos utilizar una prueba de dos colas, utilizaríamos el 2,5% superior e inferior de las 70 diferencias, es decir, sólo la diferencia superior e inferior).

3. Agrupar las observaciones de cada división posible en dos grupos (del tamaño de las agrupaciones originales). Como ya dijimos, existen procedimientos sistemáticos para realizar todas las combinaciones, pero esto es algo que en una situación real de investigación la computadora lo haría automáticamente. La tabla 15-9 muestra las 70 divisiones.

4. Calcular la diferencia media entre los grupos de cada división. Por ejemplo, en la segunda división que aparece en el ejemplo, las medias son 7,5 para el grupo "**No**" y 44,5 para el grupo "**Si**", dando una diferencia entre las medias igual a 37. Debajo de cada una de las 70 divisiones se indican las diferencias entre las medias.

5. Ordenar las diferencias de menor (más negativa) a mayor. En el ejemplo hay 70 diferencias, que van desde una diferencia de –37, donde los no altamente sensibles presentan una mayor cantidad de libros leídos, a +37, donde los altamente sensibles presentan una mayor cantidad de libros leídos.

Tabla 15-9.
Cálculos de una prueba de aleatorización basada en el estudio que compara niños altamente sensibles y no altamente sensibles en cuanto a la cantidad de libros leídos durante el año anterior (datos ficticios).

Resultados reales:

	Altamente sensibles	
	No	*Si*
	0	17
	3	36
	10	45
	22	75
Σ	35	173
$M =$	8,75	43,25

Diferencia real = $M_{Si} - M_{No} = 34{,}5$
Para rechazar la hipótesis nula: la diferencia media obtenida debe ubicarse entre el 5% superior de diferencias medias. Con 70 diferencias medias, debe estar entre las tres diferencias superiores.

Todas las divisiones posibles (70) de las ocho observaciones en dos grupos de cuatro cada uno:

	Real													
	No	*Si*	*No*	*Si*	*No*	*Si*	*No*	*Si*	*No*	*Si*	*No*	*Si*	*No*	*Si*
	0	17	0	22	0	22	0	22	0	22	0	10	0	10
	3	36	3	36	3	17	3	17	3	17	3	36	3	17
	10	45	10	45	10	45	10	36	10	36	22	45	22	45
	22	75	17	75	36	75	45	75	75	45	17	75	36	75
$M_{Si} - M_{No}$	34,5		37		27,5		23		8		31		21,5	

	No	*Si*	*No*	*Si*	*No*	*Si*	*No*	*Si*	*No*	*Si*	*No*	*Si*	*No*	*Si*
	0	10	0	10	0	10	0	10	0	10	0	10	0	10
	3	17	3	17	3	22	3	22	3	22	3	22	3	22
	22	36	22	36	17	45	17	36	17	36	36	17	36	17
	45	75	75	45	36	75	45	75	75	45	45	75	75	45
$M_{Si} - M_{No}$	17		2		24		19,5		4,5		10		–5	

	No	*Si*	*No*	*Si*	*No*	*Si*	*No*	*Si*	*No*	*Si*	*No*	*Si*	*No*	*Si*
	0	10	0	3	0	3	0	3	0	3	0	3	0	3
	3	22	10	36	10	17	10	17	10	17	10	22	10	22
	45	17	22	45	22	45	22	36	22	36	17	45	17	36
	75	36	17	75	36	75	45	75	75	45	36	75	45	75
$M_{Si} - M_{No}$	–9,5		27,5		18		13,5		–1,5		20,5		16	

	No	*Si*	*No*	*Si*	*No*	*Si*	*No*	*Si*	*No*	*Si*	*No*	*Si*	*No*	*Si*
	0	3	0	3	0	3	0	3	0	3	0	3	0	3
	10	22	10	22	10	22	10	22	22	10	22	10	22	10
	17	36	36	17	36	17	45	17	17	45	17	36	17	36
	75	45	45	75	75	45	75	36	36	75	45	75	75	45
$M_{Si} - M_{No}$	1		6,5		–8,5		–13		14,5		10		–5	

	No	*Si*	*No*	*Si*	*No*	*Si*	*No*	*Si*	*No*	*Si*	*No*	*Si*	*No*	*Si*
	0	3	0	3	0	3	0	3	0	3	0	3	0	3
	22	10	22	10	22	10	17	10	17	10	17	10	36	10
	36	17	36	17	45	17	36	22	36	22	45	22	45	22
	45	75	75	45	75	36	45	75	75	45	75	36	75	17
$M_{Si} - M_{No}$	0,5		–14,5		–19		3		–12		–16,5		–26	

Tabla 15-9 (*cont.*).

	No	*Si*	*No*	*Si*	*No*	*Si*	*No*	*Si*	*No*	*Si*	*No*	*Si*	*No*	*Si*
	17	0	22	0	22	0	22	0	22	0	10	0	10	0
	3	3	36	3	17	3	17	3	17	3	36	3	17	3
	10	10	45	10	45	10	36	10	36	10	45	22	45	22
	22	22	75	17	75	36	75	45	45	75	75	17	75	36
$M_{Si} - M_{No}$	−34,5		−37		−27,5		−23		−8		−31		−21,5	
	No	*Si*	*No*	*Si*	*No*	*Si*	*No*	*Si*	*No*	*Si*	*No*	*Si*	*No*	*Si*
	10	0	10	0	10	0	10	0	10	0	10	0	10	0
	17	3	17	3	22	3	22	3	22	3	22	3	22	3
	36	22	36	22	45	17	36	17	36	17	17	36	17	36
	75	45	45	75	75	36	75	45	45	75	75	45	45	75
$M_{Si} - M_{No}$	−17		−2		−24		−19,5		−4,5		−10		5	
	No	*Si*	*No*	*Si*	*No*	*Si*	*No*	*Si*	*No*	*Si*	*No*	*Si*	*No*	*Si*
	10	0	3	0	3	0	3	0	3	0	3	0	3	0
	22	3	36	10	17	10	17	10	17	10	22	10	22	10
	17	45	45	22	45	22	36	22	36	22	45	17	36	17
	36	75	75	17	75	36	75	45	45	75	75	36	75	45
$M_{Si} - M_{No}$	9,5		−27,5		−18		−13,5		1,5		−20,5		−16	
	No	*Si*	*No*	*Si*	*No*	*Si*	*No*	*Si*	*No*	*Si*	*No*	*Si*	*No*	*Si*
	3	0	3	0	3	0	3	0	3	0	3	0	3	0
	22	10	22	10	22	10	22	10	10	22	10	22	10	22
	36	17	17	36	17	36	17	45	45	17	36	17	36	17
	45	75	75	45	45	75	36	75	75	36	75	45	45	75
$M_{Si} - M_{No}$	−1		−6,5		8,5		13		−14,5		−10		5	
	No	*Si*	*No*	*Si*	*No*	*Si*	*No*	*Si*	*No*	*Si*	*No*	*Si*	*No*	*Si*
	3	0	3	0	3	0	3	0	3	0	3	0	3	0
	10	22	10	22	10	22	10	17	10	17	10	17	10	36
	17	36	17	36	17	45	22	36	22	36	22	45	22	45
	75	45	45	75	36	75	75	45	45	75	36	75	17	75
$M_{Si} - M_{No}$	−0,5		14,5		19		−3		12		16,5		26	

Las setenta diferencias ordenadas de menor (más negativa) a mayor:

-37, -34,5, -32, -27,5, -27,5, -26, -21,5, -24, -23, -20,5, -19,5, -16, -16,5, -17, -18, -19, -14,5, -14,5, -13,5, -13, -12, -10, -10, -9,5, -8,5, -8, -6,5, -5, -5, -4,5, -1,5, -3, -2, -1, -0,5, 0,5, 1, 1,5, 2, 3, 4,5, 5, 5, 6,5, 8, 8,5, 9,5, 10, 10, 12, 13, 13,5, 14,5, 14,5, 16, 16,5, 17, 18, 19, 19,5, 20,5, 21,5, 23, 24, 26, 27,5, 27,5, 31, $\boxed{34,5}$, 37

Conclusión: la diferencia media real se ubica entre las tres superiores. Se rechaza la hipótesis nula.

6. Comparar el punto de corte con el lugar en donde se ubica la diferencia real dentro de la lista ordenada, para determinar si se rechaza o no la hipótesis nula. En el ejemplo que analizamos, la diferencia real de +34,5 es la segunda superior, ubicándose en un lugar entre las tres superiores, tal como se requería. Podemos rechazar la hipótesis nula.

Otro ejemplo de prueba de aleatorización

Esta vez analizaremos el ejemplo de la prueba de álgebra y del nivel escolar del capítulo anterior, que incluye una correlación. Si el alumno aún no ha estudiado el capítulo 3, debería saltear esta sección.

Tabla 15-10.
Cálculos de una prueba de aleatorización basada en el estudio que correlaciona la nota de nivel escolar y la puntuación en una prueba de álgebra (datos ficticios).

Para rechazar la hipótesis nula: la correlación real debe ser la mayor de las 24 correlaciones posibles para poder rechazar la hipótesis nula al nivel del 5%, con prueba de una cola.

Correlaciones de todas las posibles combinaciones entre las OPA (Observaciones de pruebas de álgebra) y los NE (Niveles escolares).

OPA (Real)	NE (Real)	OPA	NE	OPA	NE	OPA	NE	OPA	NE
1	4	1	6	1	9	1	7	1	9
4	6	4	9	4	7	4	4	4	6
10	9	10	7	10	4	10	6	10	7
95	7	95	4	95	6	95	9	95	4
$r = 0,25$		$r = -0,79$		$r = -0,24$		$r = 0,79$		$r = -0,82$	

OPA	NE	OPA	NE	OPA	NE	OPA	NE	OPA	NE
1	7	1	9	1	7	1	4	1	4
4	6	4	4	4	6	4	7	4	9
10	9	10	6	10	4	10	6	10	7
95	4	95	7	95	9	95	9	95	6
$r = -0,76$		$r = 0,12$		$r = 0,75$		$r = 0,82$		$r = 0,52$	

OPA	NE	OPA	NE	OPA	NE	OPA	NE	OPA	NE
1	7	1	4	1	4	1	6	1	6
4	4	4	9	4	7	4	4	4	4
10	9	10	6	10	9	10	9	10	7
95	6	95	7	95	6	95	7	95	9
$r = -0,11$		$r = 0,18$		$r = -0,08$		$r = 0,22$		$r = 0,82$	

OPA	NE	OPA	NE	OPA	NE	OPA	NE	OPA	NE
1	6	1	6	1	9	1	9	1	7
4	9	4	7	4	4	4	7	4	9
10	4	10	4	10	7	10	6	10	6
95	7	95	9	95	6	95	4	95	4
$r = 0,11$		$r = 0,76$		$r = 0,18$		$r = -0,84$		$r = 0,82$	

OPA	NE	OPA	NE	OPA	NE	OPA	NE
1	4	1	6	1	9	1	7
4	6	4	9	4	7	4	9
10	7	10	4	10	6	10	4
95	9	95	7	95	4	95	6
$r = 0,84$		$r = 0,11$		$r = -0,84$		$r = -0,22$	

Correlaciones de menor a mayor:

–0,84, –0,84, –0,82, –0,79, –0,76, –0,24, –0,22, –0,18, –0,11, –0,08, 0,11, 0,11, 0,12, 0,18, 0,22, $\boxed{0,25}$ 0,52, 0,75, 0,76, 0,79, 0,82, 0,82, 0,82, 0,84

Conclusión: no se rechaza la hipótesis nula.

Una de las formas de *realizar una prueba de aleatorización* de una correlación implica calcular una correlación entre cada posible combinación de observaciones de las dos variables (sin combinar nunca dos observaciones de la misma variable). Con cuatro participantes, existen 24 combinaciones posibles de este tipo. Para que una correlación sea significativa al 5%, la correlación de la combinación real de observaciones de la muestra real debe ser la mayor de las 24 correlaciones posibles. La tabla 15-10 muestra los cálculos de la prueba de aleatorización. Utilizando este procedimiento, la correlación no resultó significativa. (Es el mismo resultado que obtuvimos anteriormente en el capítulo para el mismo ejemplo, utilizando una transformación *log*).

Prueba de aleatorización aproximada

Hemos ilustrado la prueba de aleatorización utilizando muestras muy pequeñas. Aun así, existían bastantes posibles divisiones de las observaciones en cada ejemplo. Con muestras más grandes (y más realistas), la cantidad de distintas divisiones rápidamente se hace inmanejable, incluso para la mayoría de las computadoras. Por ejemplo, una comparación entre dos grupos de siete participantes cada uno tiene 3.432 divisiones posibles; una comparación de 10 participantes por grupo tiene 184.756. ¡Con 20 por grupo, hay 155.120.000 posibles divisiones! En la práctica, aún la mayoría de las computadoras no están en condiciones de manejar verdaderas pruebas de aleatorización con los tamaños de muestra comunes en la investigación psicológica.

Para solucionar este problema, los estadísticos han desarrollado lo que se denomina **prueba de aleatorización aproximada**. La computadora selecciona al azar una gran cantidad de posibles divisiones de la muestra, tal vez 100 ó incluso 1.000. Los resultados logrados utilizando estas divisiones seleccionadas al azar se consideran, entonces, representativos de lo que encontraríamos si utilizáramos realmente cada división posible. (El proceso es similar a un estudio Montecarlo, *descripto en el cuadro* 10-1. ¿Cómo algo tan metódico como una computadora produce tantos números aleatorios?, *véase* el cuadro 15-1).

El otro método que mencionamos al comienzo, el *boot strap*, también trabaja del mismo modo: la computadora genera una gran cantidad de selecciones aleatorias, y el resultado real se compara con los resultados teóricamente posibles que ha seleccionado la computadora. Las únicas diferencias entre los procedimientos residen en los tipos de divisiones o combinaciones de observaciones que se seleccionan aleatoriamente; pero estos son asuntos técnicos que exceden el alcance de un libro de nivel introductorio.

COMPARACIÓN DE MÉTODOS

Hemos analizado tres métodos para realizar pruebas de hipótesis cuando las muestras parecen provenir de poblaciones no normales: transformación de datos, pruebas de rango y orden y métodos intensivos por computadora, tales como las pruebas de aleatorización. ¿Cómo decide un investigador el método a utilizar?

Las *transformaciones de datos* tienen la ventaja de permitir aplicar las técnicas paramétricas familiares a los valores transformados. Pero las transformaciones no siempre funcionan. Es decir, puede no existir ninguna transformación razonable que produzca valores normales en todos los grupos. Además, las transformaciones pueden distorsionar las observaciones de modo que se pierda el significado original.

Los métodos de rango y orden pueden aplicarse independientemente de las distribuciones. Son especialmente adecuados cuando las observaciones originales son rangos, y también son útiles cuando las observaciones no siguen claramente un patrón numérico simple (medición intervalar), situación que algunos psicólogos consideran bastante común. Más aún, la lógica de los métodos de rango y orden es simple y directa, y no requiere construcciones elaboradas de distribuciones hipotéticas o parámetros estimados.

Sin embargo, los métodos de rango y orden no son tan familiares para aquellos que leen publicaciones científicas, y tampoco han sido desarrollados para muchas situaciones complejas. Otro problema es que la lógica simple de las pruebas de rango y orden se pierde si existen muchos rangos iguales. Finalmente, al igual que los métodos de transformación de datos, los métodos de rango y orden distorsionan los datos originales, perdiéndose información. Por ejemplo, en la misma muestra, una diferencia entre 6,1 y 6,2 podría ser un rango, pero la diferencia entre 3,4 y 5,8 también podría ser un rango.[6]

Los métodos intensivos por computadoras, tales como las pruebas de aleatorización aproximada, no requieren ninguno de los dos supuestos principales de las pruebas paramétricas ordinarias. Más aún, al igual que las pruebas de rango y orden, tienen una lógica directa propia que es muy atractiva, evitando todo el proceso de construcción de distribuciones estimadas de población, distribuciones de medias, etc. Los métodos intensivos por computadora son también extremadamente flexibles. Se los puede utilizar en casi cualquier situación imaginable en la que pudiera aplicarse una prueba de hipótesis. Por lo tanto, frecuentemente pueden utilizarse cuando no existen otros tipos de pruebas disponibles, paramétricas o de cualquier otro tipo.

La principal desventaja de los métodos intensivos por computadora es que son bastante nuevos; por lo tanto, los detalles y ventajas relativas de varios de los métodos no han sido bien aprovechados. Más aún, por ser nuevos, en la mayoría de los casos los paquetes estadísticos estándar para computadoras no los incluyen. Los métodos intensivos por computadora recién están empezando a aparecer en las publicaciones científicas, pero es probable que su aplicación aumente con rapidez.

Riesgo relativo de cometer errores Tipo I y Tipo II

¿Cuál es la precisión de los distintos métodos en cuanto a que el nivel del 5% realmente implica que existe un 5% de probabilidad de rechazar incorrectamente la hipótesis nula? y ¿cómo afectan la potencia los distintos métodos?

Cuando se cumplen los supuestos de las pruebas paramétricas, estas pruebas son tan buenas o mejores que cualquiera de las alternativas. Lo expresado en el párrafo anterior es cierto en cuanto a la protección contra los errores Tipo I y Tipo II, situación que era de esperarse, ya que se dan las condiciones para las cuales fueron diseñadas las pruebas paramétricas.

Sin embargo, cuando no se cumplen los supuestos de las pruebas paramétricas, las ventajas relativas de los tres posibles procedimientos alternativos que hemos analizado no son del todo claras. De hecho, los méritos relativos de los distintos procedimientos son temas de activa controversia, sobre los cuales se publican muchos artículos cada año en las revistas especializadas en estadística.

La razón de la controversia es que el procedimiento más adecuado depende de los tipos de distribuciones involucradas. Una distribución no normal puede serlo de muchas maneras (*véase* capítulo 5). Sucede que los efectos de los diferentes métodos sobre los errores Tipo I y Tipo II va-

[6] Otra ventaja tradicional de las pruebas de rango y orden ha sido su facilidad de cálculo. Excepto por el trabajo de convertir las observaciones en rangos, los cálculos reales de la mayoría de estos procedimientos son muy simples, comparados con los de las pruebas paramétricas. Actualmente, con la utilización de las computadoras, es igualmente fácil calcular cualquier tipo de procedimiento. Con algunos paquetes estadísticos estándar para computadoras, es realmente mucho menos problemático calcular las pruebas paramétricas. Además, a veces la prueba de rango y orden apropiada puede no estar disponible.

Cuadro 15-1.
¿De dónde provienen los números aleatorios?

Para ser aleatorios, los números deben ser obtenidos teniendo todos la misms probabilidad de ser seleccionados. Es decir, la posibilidad de que surja cualquier número debe ser totalmente independiente de las posibilidades del número que surja con anterioridad o posterioridad a él. Una de las muchas aplicaciones importantes de los números aleatorios son los métodos estadísticos intensivos por computadora, tal como lo hemos visto en este capítulo. También son fundamentales para los estudios Montecarlo (*véase* cuadro 10-1), estudios que se utilizan para probar el efecto del incumplimiento de la normalidad y otros supuestos de las pruebas estadísticas paramétricas, y que constituyen uno de los medios con los que cuentan los psicólogos para saber si necesitan utilizar los métodos descriptos en este capítulo. Sin embargo, los números aleatorios son, en sí mismos, un tema interesante.

La primera tabla de números aleatorios se creó en 1927. Con anterioridad a esa fecha, se utilizaban métodos mecánicos tales como dispositivos para mezclar. El alumno seguramente recordará a William S. ("*Student*") Gosset (cuadro 9-1). Para obtener sus números aleatorios, Gosset mezcló y extrajo números de un mazo de 3.000 cartas. Luego, en 1927, Karl Pearson incentivó a L. H. C. Tippett para que publicara cierta tabla. Tippett consideraba que extraer cartas numeradas de una bolsa era "insatisfactorio"; por eso seleccionó dígitos del censo de 1925. Más tarde, en 1938, R. A. Fisher y Frank Yates publicaron una lista basada en logaritmos. Casi al mismo tiempo, también fue presentada una cantidad de métodos de control de aleatoriedad.

Más tarde, se hicieron comunes soluciones físicas más sofisticadas. Una de ellas consistía en hacer brillar un rayo de luz a intervalos regulares sobre un disco giratorio dividido en secciones. Otro método utilizaba la radiación de sustancias radioactivas: registraba la cantidad de partículas detectadas durante cierto periodo; si la cantidad era impar, establecía el contador en 1; si era par, en 0, y luego generaba listas de números a partir de agrupaciones de esos dígitos binarios. Un tercer sistema empleaba una válvula electrónica que emitía un sonido que podía ser amplificado; los valores fluctuantes de la potencia de salida eran valores aleatorios.

Todos estos métodos físicos eran una incomodidad: era necesario guardar los números si iban a ser reproducidos o utilizados nuevamente, y todos los aparatos utilizados eran difíciles de mantener. Por eso, en la actualidad, con frecuencia se utilizan computadoras para crear "números seudoaleatorios", utilizando alguna ecuación especial, como elevar grandes números al cuadrado y tomar un grupo central de los dígitos resultantes. Pero estos números, en un sentido muy sutil, no son aleatorios sino predecibles por el propio hecho de que había una intención en el diseño de la ecuación: crear azar (vaya paradoja). También existe el inconveniente de que las ecuaciones puedan "degenerarse" y comenzar a repetir secuencias. Finalmente, no importa cómo se genere la lista. Existe controversia acerca de las consecuencias de la utilización reiterada de la misma tabla.

El tema de la dificultad de crear algo libre de orden o inteligencia parece estar indicando algo. Dejaremos que el alumno lo decida.

rían según el tipo de distribución de que se trate. Incluso, para determinado tipo de distribución, una técnica podría resultar mejor cuando los grupos tienen las mismas cantidades y otra cuando las cantidades en cada grupo son distintas; o bien, un método podría ser mejor con un gran tamaño de muestra y otro con una muestra pequeña. Más aún, al comparar grupos, las distribuciones de los grupos pueden incluir diferentes tipos de distribuciones no normales.

Aunque se han realizado muchos estudios comparando los distintos métodos (*véase* cuadro 10-1 acerca de los estudios Montecarlo), aún sabemos muy poco sobre la efectividad relativa de estos métodos en la mayoría de los casos. Aún peor, en muchas situaciones, un investigador puede tener la noción de que una muestra no proviene de una población normal, pero no de qué tipo particular de población no normal se trata. Por lo tanto, incluso los estudios que se han realizado comparando los distintos procedimientos con determinadas formas de población no normal, pueden no ser demasiado útiles al momento de enfrentar los resultados de un estudio real.

Es posible que algún día se realice la suficiente investigación que abarque las suficientes cantidades de situaciones como para que surjan patrones que nos den pautas prácticas adecuadas. Por el momento, según nuestra opinión, los investigadores deben confiar en otros criterios (como los presentados en este capítulo) para seleccionar entre las distintas alternativas, cuando los supuestos no se han cumplido. Sin embargo, desde el punto de vista de la lectura de investigaciones (un tema que trataremos a continuación), lo que necesitamos es poder comprender la lógica del procedimiento en particular que ha elegido el investigador. Decidir si fue elegido correctamente, tal vez sea una tarea que exceda los conocimientos del alumno en esta instancia, razón por la que puede relajarse hasta los próximos cursos y futuros avances en el área.

CONTROVERSIAS

Todos los temas tratados en este capítulo son controvertidos, especialmente, la conveniencia de las transformaciones de datos, los riesgos de utilizar procedimientos paramétricos cuando se desconocen las distribuciones poblacionales, hasta qué punto es apropiado tratar a las medidas típicas en psicología como si produjeran mediciones de intervalares y las ventajas y desventajas de los métodos intensivos por computadora. (Judd et al. 1995 nos ofrecen una revisión reciente de las controversias).

PROCEDIMIENTOS UTILIZADOS CUANDO LAS POBLACIONES NO PARECEN NORMALES, SEGÚN SE DESCRIBEN EN LAS PUBLICACIONES CIENTÍFICAS

La utilización de los procedimientos que hemos descripto en este capítulo parece tener altibajos de popularidad en las diferentes áreas de la psicología. En algunas áreas, durante ciertos años podemos encontrar muchos estudios que utilizan transformaciones de datos y no ver nunca una prueba de rango y orden. En otras áreas, podemos encontrar exactamente lo contrario. Y los métodos intensivos por computadora aplicados a la psicología son un desarrollo tan nuevo que probablemente podamos encontrarlos sólo en unos pocos de los más recientes estudios, con frecuencia en circunstancias en las que no existe ningún procedimiento alternativo evidente.

Las transformaciones de datos se mencionan comúnmente justo antes de la descripción del análisis que utiliza los valores transformados. Por ejemplo, Connors et al. (1997) realizaron un estudio concentrándose en la alianza entre el paciente y el terapeuta en los tratamientos de alcoholismo. Antes de informar los resultados de su estudio, comentaron lo siguiente:

> Las variables tales como el porcentaje de días de abstinencia y tragos por día, con frecuencia se alejan de la normalidad debido a la asimetría y a los efectos techo y piso. En respuesta a esto, la variable del porcentaje de días de abstinencia fue sometida a una transformación arco-seno, y la variable de los tragos por día fue sometida a una transformación raíz cuadrada; en cada caso, el procedimiento se realizó para mejorar la distribución (p. 592).

A continuación, presentamos un ejemplo de una prueba de rango y orden en un estudio de investigación realizado por Ford et al. (1997), que se concentró en la relación de ciertos factores de la personalidad con el tratamiento de trastornos de estrés postraumático (una condición psicológica que resulta de un hecho traumático como el que podría ser experimentado durante una guerra o como resultado de un ataque violento). El factor de personalidad de interés para los investigadores se basaba en una versión moderna de la teoría psicoanalítica freudiana denominada "relaciones objeto", que se refiere al impacto psicológico de nuestras primeras relaciones, principalmente con nuestros padres (los "objetos" de estas relaciones tempranas). Los investigadores basaron sus medidas sobre las relaciones objeto en una entrevista clínica concentrada en temas tales como la capacidad de dedicarse a una relación estrecha y la capacidad de tener una imagen compleja de los otros (p. ej. no ver a una persona como completamente buena o completamente mala). Al informar sus resultados, abreviaron la medida de entrevista clínica como RO-C (Relaciones objeto). La distribución de registros de las RO-C no era normal (era bimodal).

Uno de sus análisis se concentraba en la asociación de las relaciones objeto con el hecho de que una persona continúe con el tratamiento hasta completarlo o lo interrumpa prematuramente. Informaron sus resultados de la siguiente manera:

> Seis de los 74 participantes interrumpieron prematuramente el tratamiento [...] Los seis pacientes que interrumpieron prematuramente no difieren del resto de la muestra en cuanto a ninguna variable demográfica o de prueba preliminar [...] Sí difieren en forma estadísticamente significativa de aquellos que completaron el tratamiento en cuanto a las RO-C, con valores menores según la prueba *U* de Mann-Whitney no paramétrica ($Z = -3{,}43$, $p < 0{,}001$) (p. 554).

Se puede observar que informa un valor Z, lo que sugiere que en lugar de utilizar la prueba *U* de Mann-Whitney directa utilizaron el procedimiento de aproximación normal.

Finalmente, un estudio realizado por Caspi y Herbener (1990) nos ofrece un ejemplo del método intensivo por computadora informado en una publicación científica. Como parte del estudio, los investigadores analizaron la estabilidad a largo plazo de la personalidad de 252 individuos que fueron probados primero en el año 1970 y luego nuevamente en el año 1981. En cada prueba, los participantes completaron la prueba de personalidad denominada "*Q sort*". Se trata de un tipo especial de procedimiento de prueba en el que el participante recibe cierta cantidad de cartas, cada una con un rasgo de personalidad. El participante luego pone estas cartas en pilones, clasificándolas desde "**para nada descriptiva**" hasta "**altamente descriptiva**". Lo particular del método, sin embargo, es que el participante debe ubicar las cartas en pilones de determinados tamaños, tamaños que corresponden a una curva normal, con más cartas en el medio y menos en los extremos.

Para estudiar la estabilidad, Caspi y Herbener tuvieron que correlacionar los *Q sort* de los dos periodos. Sin embargo, los autores observaron que las correlaciones entre los *Q sort* ("correlaciones *Q*"), del modo en que se utilizaban en su estudio, tenían algunas propiedades estadísticas inusuales. Los autores explicaron:

> Realizamos lo que esencialmente sería una prueba de aleatorización. Específicamente, el perfil *Q sort* de cada sujeto en el año 1970 fue correlacionado con su perfil del *Q sort* en 1981 para todos los sujetos del mismo sexo [...] Para cada sexo [...] generamos 100 muestras aleatorias para comparar con las correlaciones *Q* reales indicadoras de la estabilidad de la personalidad [...] En el caso de las [...] mujeres, las 100 pruebas no arrojaron valores que excedieran la media muestral [de correlaciones de los perfiles de cada sujeto en 1970 y 1981] (0,49) (p. 253).

En cuanto a los hombres, los resultados fueron similares.

Resumen

La prueba t, el análisis de varianza y la prueba de significación del coeficiente de correlación suponen que las poblaciones siguen una distribución normal. Cuando las muestras sugieren que las poblaciones están muy lejos de lo normal (por ejemplo, debido a casos atípicos), utilizar los procedimientos ordinarios arroja resultados incorrectos.

Un método a utilizar, cuando la población parece no ser normal, es transformar los valores, como por ejemplo sacando la raíz cuadrada de cada valor para que la distribución de los valores transformados aparente representar una población normalmente distribuida. Otras transformaciones comunes para distribuciones asimétricas se realizan calculando el logaritmo de cada valor o su inverso. Después de realizadas las transformaciones pueden aplicarse los procedimientos ordinarios de prueba de hipótesis.

Otro método para resolver el problema es ordenar todas las observaciones del estudio por rango. Las pruebas especiales de rango y orden (a veces denominadas no paramétricas o libres de distribución) utilizan principios básicos de probabilidad para determinar la posibilidad de que los rangos estén irregularmente distribuidos en el grupo experimental.

Una prueba de aleatorización es un ejemplo del método intensivo por computadora, que analiza cada posible disposición de las observaciones de un estudio para determinar la probabilidad de que la disposición obtenida (en términos, por ejemplo, de la diferencia de medias entre los grupos) surja por casualidad. Sin embargo, dado que incluso con computadoras las pruebas de aleatorización no son prácticas con muestras de tamaños razonables, se utilizan otros métodos intensivos por computadora. Por ejemplo, 1.000 de las posibles disposiciones son seleccionados al azar y la distribución de sus diferencias de medias se compara con la obtenida en la muestra real.

Las transformaciones de datos permiten utilizar técnicas paramétricas que resultan familiares, pero no siempre pueden ser aplicadas y pueden distorsionar el significado de los datos. Los métodos de rango y orden pueden aplicarse a muchas series de datos; son especialmente adecuados con rangos o datos similares, y tienen una base conceptual directa. Pero las técnicas de rango y orden no son ampliamente conocidas y no han sido desarrolladas para muchas situaciones complejas de análisis de información. Al igual que con otras transformaciones de datos, la información puede perderse o el significado distorsionarse. Los métodos intensivos por computadora son ampliamente aplicables, a veces incluso en situaciones para las cuales no existe otro método disponible. Además, tienen una atrayente lógica básica. Pero a los investigadores no les resultan muy familiares; al ser nuevos, sus posibles limitaciones no están bien resueltas; y pueden ser difíciles de emprender ya que no se incluyen en programas estándar para computadoras. Cuando se supone que la población no es normal, no existe demasiado acuerdo acerca de cuál de las técnicas es más conveniente en cuanto al riesgo relativo de cometer los errores Tipo I y Tipo II.

Las publicaciones científicas generalmente describen las transformaciones de datos justo antes del análisis que las utiliza. Los métodos de rango y orden se describen, mayormente, como cualquier otro tipo de prueba de hipótesis. Los métodos intensivos para computadoras, por ser menos conocidos, por lo general se describen con bastante detalle.

Términos clave

- Pruebas de aleatorización aproximada.
- Transformación de datos.
- Pruebas libres de distribución.
- Transformación inversa.
- Transformación *log*.
- Pruebas no paramétricas.
- Pruebas paramétricas.
- Pruebas de aleatorización.
- Pruebas de rango y orden.
- Transformación de rango y orden.
- Reflejar.
- Transformación raíz cuadrada.

Ejercicios

Los ejercicios implican la realización de cálculos (con la ayuda de una calculadora). La mayoría de los problemas estadísticos reales se resuelven por computadora, pero aunque exista la posibilidad de utilizarla, es conveniente realizar estos ejercicios manualmente para incorporar el método de trabajo.

Para adquirir práctica en la utilización de una computadora, para resolver problemas estadísticos, se puede utilizar la sección de computación de cada capítulo, publicada en la *Guía de estudio y libro de tareas de computación para el alumno [Student's Study Guide and Computer Workbook]* que acompaña este libro.

Todos los datos de esta sección son ficticios (a menos que se especifique lo contrario)

Las respuestas a los ejercicios de la serie I se encuentran al final del libro.

SERIE I

1. Para la distribución de los 30 valores que aparecen abajo, a) trace un histograma (basado en frecuencias agrupadas) de los valores tal como aparecen; b) realice una transformación raíz cuadrada y un histograma (de frecuencias agrupadas) de los valores transformados, y c) convierta los valores originales en rangos y trace un histograma (agrupado) de los mismos.

9, 28, 4, 16, 0, 7, 25, 1, 4, 10, 4, 2, 1, 9, 16, 11, 12, 1, 18, 2, 5, 10, 3, 17, 6, 4, 2, 23, 21, 20

2. ¿Cuáles de las siguientes distribuciones muestrales sugieren que la distribución poblacional probablemente no es normal? Explique por qué.

a) 41, 52, 74, 107, 617
b) 221, 228, 241, 503, 511, 521
c) 0,2, 0,3, 0,5, 0,6, 0,7, 0,9, 0,11
d) −6, −5, −3, 10
e) 11, 20, 32, 41, 49, 62

3. Un investigador compara el tamaño típico de familia en 10 culturas, 5 del grupo idiomático A y 5 del grupo idiomático B. Los números correspondientes a las culturas del grupo A son 1,2, 2,5, 4,3, 3,8 y 7,2. Los números correspondientes a las culturas del grupo B son 2,1, 9,2, 5,7, 6,7 y 4,8. Sobre la base de estas 10 culturas, ¿difiere el tamaño típico de la familia en las culturas de diferentes grupos idiomáticos? Utilice el nivel 0,05. a) Realice una transformación raíz cuadrada (para simplificar las cosas, redondee los valores transformados para que tengan un sólo decimal). b) Realice una prueba t para medias independientes utilizando los valores transformados (muestre su trabajo). c) Explique lo que ha hecho y por qué a una persona que está familiarizada con la prueba t pero no con la transformación de datos.

4. Un investigador asigna participantes al azar para que observen uno de tres tipos de películas: una tiende a entristecer a las personas, otra tiende a alegrar a las personas y una tercera tiende a poner furiosas a las personas. Des-

pués se pide a los participantes que califiquen unas series de fotos de individuos en cuanto al nivel de honestidad que reflejan. Las calificaciones del grupo que vio la película que causa tristeza fueron 201, 523 y 614; las calificaciones del grupo que vio la película que causa enojo fueron 136, 340 y 301 y las calificaciones del grupo que vio la película que causa alegría fueron 838, 911 y 1.007. a) Transforme las observaciones en rangos. b) Realice un análisis de varianza de un criterio con los valores transformados en rangos (utilice el nivel 0,05 de significación y muestre su trabajo). c) Explique lo que ha hecho y por qué a una persona que comprende el análisis de varianza pero no las transformaciones a rango o las pruebas no paramétricas.

5. Un estudio compara el rendimiento de personas en la realización de una tarea original: si la realizan a solas o en presencia de un amigo. Los valores correspondientes a los participantes que están solos son 9, 5 y 4; los valores correspondientes a los participantes que realizan la tarea frente a un amigo son 3, 1 y 0. a) Realice una prueba de aleatorización comparando los dos grupos. (Utilice $p < 0{,}05$, una cola, prediciendo mayores valores para los que realizan la tarea a solas). b) Explique lo que hizo a una persona que nunca ha asistido a un curso de estadística.

Nota: con tres participantes en cada grupo, existen 20 formas diferentes de realizar dos agrupaciones con los seis registros:

9 3	9 4	9 4	9 4	9 5	9 5	9 5	9 5	9 5	9 5
5 1	5 1	5 3	5 3	4 1	4 3	4 3	3 4	3 4	1 4
4 0	3 0	1 0	0 1	3 0	1 0	0 1	1 0	0 1	0 3

3 9	4 9	4 9	4 9	5 9	5 9	5 9	5 9	5 9	5 9
1 5	1 5	3 5	3 5	1 4	3 4	3 4	4 3	4 3	4 1
0 4	0 3	0 1	1 0	0 3	0 1	1 0	0 1	1 0	3 0

6. Miller (1997) realizó un estudio acerca del compromiso en una relación amorosa y la atención prestada a alternativas atractivas. En el estudio, se le mostró a los participantes una serie de diapositivas de personas atractivas. Al comienzo de la sección de Resultados, Miller observa:

> Los informes propios indicados en el Índice de atención a alternativas y el tiempo efectivamente transcurrido analizando las diapositivas atractivas del sexo opuesto [...] eran positivamente asimétricos; por lo tanto, los datos fueron transformados a logaritmos (p. 760).

Explique lo que aquí se describe (y por qué se realiza) a una persona que comprende la estadística paramétrica ordinaria pero que nunca a escuchado hablar de transformaciones de datos.

SERIE II

1. Con la distribución de 20 valores que aquí presentamos a) realice un histograma (basado en frecuencias agrupadas) de los valores tal cual los presentamos; b) realice una transformación *log* y un histograma (de frecuencias agrupadas) de los valores transformados, y c) transforme los valores originales en rangos y realice un histograma (agrupado) de los mismos. (**Nota**: para realizar la transformación *log* utilice una calculadora con la función *log* para calcular logaritmos o una computadora).

2, 207, 894, 107, 11, 79, 112, 938, 791, 3, 13, 89, 1.004, 92, 1.016, 107, 87, 91, 870, 921

2. ¿Cuál de las siguientes distribuciones muestrales sugiere que la distribución poblacional probablemente no es normal? ¿Por qué?

a) 281, 283, 287, 289, 291, 300, 302
b) 1, 4, 6, 6, 7, 7, 9, 13
c) 7, 104, 104, 104, 1.245, 1.247, 1.248, 1.251
d) 68, 74, 76, 1,938
e) 407,2, 407,5, 407,6, 407,9

3. Un psicólogo realiza un estudio a seis electricistas desempleados, correlacionando la cantidad de semanas sin empleo con la satisfacción marital. Los resultados aparecen abajo. a) Realice un diagrama de dispersión y calcule la correlación entre los valores dados. b) Realice una transformación a la raíz cuadrada de los valores correspondientes a las semanas sin em-

pleo. c) Realice un diagrama de dispersión y calcule la correlación utilizando los valores transformados. d) Compare los resultados de los dos métodos. (**Nota:** el ejercicio supone que el alumno ya ha estudiado el capítulo 3).

Semanas desempleado	Satisfacción marital
2	8
1	9
9	6
16	3
25	5
4	7

4. Un investigador realizó un experimento organizado en torno a un importante discurso televisado del presidente de los EE.UU. Inmediatamente después del discurso, tres participantes fueron asignados al azar para escuchar los comentarios del comentarista político del canal de televisión. A otros tres se les asignó pasar el mismo tiempo con la televisión apagada, reflexionando tranquilamente sobre el discurso. Después, los participantes de ambos grupos completaron un cuestionario que evaluaba cuánto del contenido del discurso recordaban con precisión. El grupo que escuchó a los comentaristas presentó los valores 4, 0 y 1. El grupo que reflexionó tranquilamente presentó los valores 9, 3 y 8.

Escuchar a los comentaristas ¿afectó el recuerdo del discurso? Utilice el nivel 0,05, una cola, prediciendo mayores valores para el grupo que reflexionó tranquilamente sobre el discurso. a) Realice una prueba *t* para medias independientes. b) Realice una prueba de aleatorización con los datos. c) Compare los resultados utilizando los dos métodos. d) Explique lo que ha hecho y los resultados a alguien que está familiarizado con la prueba *t* pero no con pruebas de aleatorización.

Nota: las 20 formas diferentes de formar dos agrupaciones con estos seis valores son las siguientes:

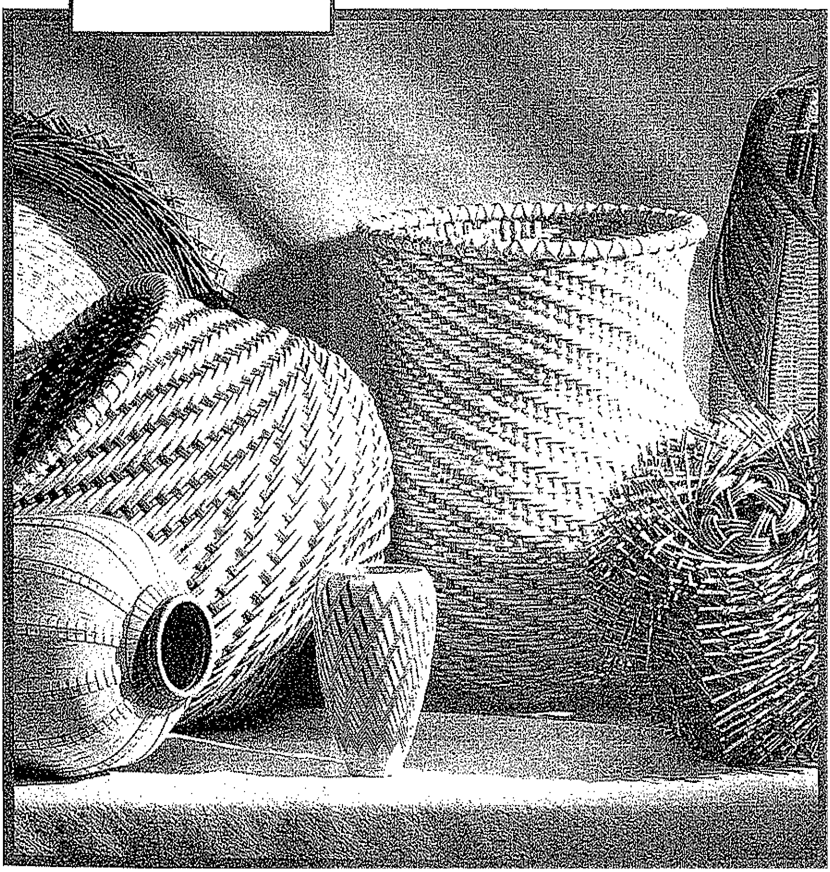
4 9	4 1	4 1	4 1	4 0	4 0	4 0	4 0	4 0	4 0
0 3	0 3	0 9	0 9	1 3	1 9	1 9	9 1	9 1	3 1
<u>1 8</u>	<u>9 8</u>	<u>3 8</u>	<u>8 3</u>	<u>9 8</u>	<u>3 8</u>	<u>8 3</u>	<u>3 8</u>	<u>8 3</u>	<u>8 9</u>
9 4	1 4	1 4	1 4	0 4	0 4	0 4	0 4	0 4	0 4
3 0	3 0	9 0	9 0	3 1	9 1	9 1	1 9	1 9	1 3
<u>8 1</u>	<u>8 9</u>	<u>8 3</u>	<u>3 8</u>	<u>8 9</u>	<u>8 3</u>	<u>3 8</u>	<u>8 3</u>	<u>3 8</u>	<u>9 8</u>

5. Un estudió comparó a los alumnos del primer y segundo año de facultad en cuanto a la cantidad de amigos íntimos. Los investigadores predijeron que los alumnos de segundo año tendrían más amigos íntimos. Los cinco alumnos de primer año que participaron de la prueba informaron 2, 0, 2, 1 y 1. Los cinco alumnos de segundo año que participaron de la prueba informaron 3, 4, 1, 2 y 6. Realice una prueba de aleatorización aproximada con estos datos de la siguiente manera: a) Calcule la diferencia de medias entre los dos grupos reales. b) Escriba la cantidad de amigos íntimos para cada participante en una tarjeta. c) Mezcle las 10 tarjetas y colóquelas boca arriba en dos grupos de cinco. Calcule la media de las primeras cinco y calcule la diferencia con la media de las segundas cinco, y luego anótela. d) Vuelva a mezclar y repita ese proceso 40 veces. e) Determine cuántas de las 40 veces obtuvo diferencias de medias tan altas como la muestra real.

6. Carey et al. (1997) desarrollaron un programa diseñado para mejorar la motivación con el fin de evitar los riesgos de infección con HIV. Después, analizaron su efectividad con un grupo de mujeres de ciudad de baja posición económica que fueron asignadas al azar para recibir el programa o formar parte del grupo de control. Todas las mujeres fueron medidas antes, 3 semanas después y 12 semanas después de que el grupo experimental participara del programa. Una de las medidas aplicadas en el estudio se refería a la comunicación sexual, como por ejemplo, hasta qué punto las mujeres, según lo que ellas informaban, habían hablado con sus parejas sobre sexo seguro y pruebas de HIV. Antes de describir los análisis sobre esta variable, Carey et al. observaron lo siguiente: "Los valores de comunicación eran positivamente asimétricos en las tres ocasiones; transformaciones *log* 10 $(x + 1)$ fueron la mejor corrección para lograr normalidad y fueron

utilizadas en análisis sucesivos" (p.536). Explique qué es lo que se describe aquí (y por qué se realiza) a una persona que comprende la estadística paramétrica ordinaria pero que nunca ha oído hablar sobre transformaciones de datos. (Puede ignorar la parte del "$x + 1$". A título informativo le explicamos que los investigadores sumaron un 1 a cada valor antes de realizar la transformación *log* porque algunos valores de comunicación eran igual a 0, y no se puede calcular el logaritmo de 0).

16 Integración de contenidos: el modelo lineal general



Descripción del capítulo

- Breve revisión de correlación y regresión múltiples.
- Relación entre los principales métodos estadísticos.
- Revisión de los principios de la regresión y la correlación múltiples.
- Introducción al modelo lineal general.
- El modelo lineal general y la regresión / correlación múltiples.
- Regresión y correlación bivariadas como casos especiales de regresión / correlación múltiple.
- La prueba *t* como caso especial del análisis de varianza.
- La prueba *t* como caso especial de la prueba de significación del coeficiente de correlación.
- El análisis de varianza como caso especial de la prueba de significación del coeficiente de correlación múltiple.
- Elección de pruebas estadísticas.
- Los supuestos y el modelo lineal general.
- Controversias y limitaciones.
- Resumen.
- Términos clave.
- Ejercicios.

El propósito de este capítulo es unificar y profundizar los conocimientos acerca de las principales técnicas aprendidas: el análisis de varianza, la prueba *t*, la correlación y la regresión; a la vez que constituye una revisión completa de las técnicas mencionadas.

RELACIÓN ENTRE LOS PRINCIPALES MÉTODOS ESTADÍSTICOS

Más del 90% de los estudios publicados en 1988 por las más prestigiosas revistas especializadas en psicología social empleaban pruebas *t*, análisis de varianza, correlación o regresión múltiples (Reis & Stiller, 1992). Probablemente, ya se hayan hecho evidentes muchas semejanzas entre estos cuatro métodos y las otras técnicas estadísticas aprendidas a lo largo del libro. De hecho, las técnicas están más relacionadas de lo que podríamos creer: muchas de ellas no son más que simples variaciones matemáticamente equivalentes entre sí, y la mayoría tienen su origen en la misma fórmula general. Lo anterior se debe a que existe una lógica central que sustenta todos estos métodos. La lógica central se basa en una fórmula general que los estadísticos matemáticos denominan **modelo lineal general.** (El modelo lineal general no tiene ninguna relación especial con el modelo estructural del análisis de varianza).

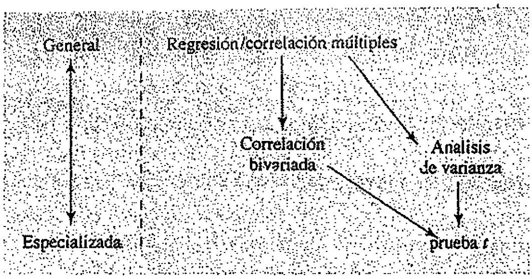


Figura 16-1.
Relación entre las cuatro principales técnicas estadísticas.

Por lo tanto, vamos a concentrarnos en los cuatro grandes, todos ellos casos especiales del modelo lineal general y, por ello, sistemáticamente relacionados. Es posible que en el proceso emerjan muchas de las intuiciones que habíamos percibido parcialmente con respecto a lo aprendido.

Para expresarlo en forma breve (y luego profundizar sobre el tema), la técnica más general es la regresión/correlación múltiples (capítulo 4), siendo la correlación bivariada (capítulo 3) un caso especial de la misma. Finalmente, la prueba *t* (capítulos 9 y 10) deriva directamente de la correlación bivariada o del análisis de varianza. La figura 16-1 representa gráficamente las vinculaciones mencionadas.

Cuando decimos que un procedimiento es un caso especial de otro, queremos decir que el primero puede deducirse de la fórmula del segundo. Por eso, cuando utilizamos los procedimientos más especializados obtenemos el mismo resultado que hubiéramos obtenido con el procedimiento más general. Para ser más concretos, si viajáramos a una isla desierta a realizar una investigación psicológica y sólo pudiéramos llevar un programa de computación para realizar las pruebas estadísticas, nos convendría elegir uno que realizara correlación/regresión múltiples. Con ese programa podríamos lograr todo lo que se obtiene con programas más especializados de correlación bivariada, pruebas *t* y análisis de varianza.

En este capítulo investigamos tales vínculos. Primero, revemos brevemente la idea de regresión/correlación múltiples que presentamos en el capítulo 4, y en ese contexto analizamos una definición formal del modelo lineal general. Después, examinamos cada uno de los vínculos: la regresión / correlación múltiples con la correlación bivariada; el análisis de varianza con la prueba *t*, y la regresión / correlación múltiples con el análisis de varianza.

REVISIÓN DE LOS PRINCIPIOS DE LA REGRESIÓN Y LA CORRELACIÓN MÚLTIPLES

Repasemos brevemente el principio de regresión y correlación vistos en los capítulos 3 y 4. En primer lugar, recordemos la idea básica de predicción bivariada (también denominada regresión bivariada) que implica formular una regla sistemática para predecir el valor de una persona en cuanto a determinada variable dependiente, a través del análisis del valor de esa persona en la variable de predicción (o independiente). Por ejemplo, predijimos los niveles de estrés de nuevos gerentes, a partir del conocimiento de la cantidad de personas que supervisarían. La regresión múltiple es aquella situación en la que se predice basándose en dos o más variables de predicción; por ejemplo, predecir el nivel de estrés de nuevos gerentes utilizando la cantidad a supervisar más el nivel de ruido y la cantidad de decisiones a tomar por mes.

Podemos crear normas de predicción tanto con puntuaciones Z como con puntuaciones originales. En este capítulo, nos concentramos en la opción de las puntuaciones originales, ya que a través de ellas se observa con mayor facilidad la relación con el modelo lineal general. Una regla de predicción para la regresión múltiple con tres variables de predicción, trabajando con puntuaciones originales, sería la siguiente: el valor a predecir para una persona con respecto a la variable dependiente es la suma de un número en particular (la constante de regresión, denominada a), más un coeficiente de regresión para puntuaciones originales (b_1) por el valor de la persona en cuestión en la primera variable de predicción (X_1); más un segundo coeficiente de regresión para puntuaciones originales (b_2) por el registro de la persona en la segunda variable de predicción (X_2), más un tercer coeficiente de regresión para puntuaciones originales (b_3) por el valor de la persona en la tercera variable de predicción (X_3).

La fórmula es la siguiente:

$$\hat{Y} = a + (b_1)(X_1) + (b_2)(X_2) + (b_3)(X_3) \tag{16-1}$$

donde $\hat{Y}$ es el valor predicho de la variable dependiente.

Por ejemplo, en el caso del nivel de estrés de los gerentes, presentado en el capítulo 4, sugerimos que una posible regla de predicción de puntuaciones originales, para un caso con tres variables de predicción, podría ser la siguiente:

$$\begin{aligned}\widehat{\text{Estrés}} = {} & -4{,}70 + (0{,}56 \text{ cantidad de personas supervisadas}) \\ & + (0{,}06 \text{ ruido medido en decibeles}) \\ & + (0{,}86 \text{ cantidad de plazos a cumplir por mes})\end{aligned}$$

Así, si un presunto gerente fuera a supervisar sólo a cuatro personas en un área con 50 decibeles de ruido, y tuviera sólo un plazo a cumplir por mes, el nivel de estrés predicho sería calculado de la siguiente manera:

$$\begin{aligned}\widehat{\text{Estrés}} &= -4{,}70 + (0{,}56)(4) + (0{,}06)(50) + (0{,}86)(1) \\ &= 4{,}70 + 2{,}24 + 3 + 0{,}86 = 1{,}40\end{aligned}$$

La predicción del nivel de estrés del gerente sería muy baja (1,40).

También es posible describir el grado general de relación entre la variable dependiente y la combinación de variables de predicción. Este dato se denomina coeficiente de correlación múltiple y se simboliza con una R. R debe ser al menos tan grande como la correlación bivariable más pequeña entre cualquiera de las variables de predicción y la variable dependiente. R^2 es la reducción proporcional del error cuadrático lograda utilizando la regla de predicción para regresión múltiple, en contraposición con la simple predicción de la variable dependiente a partir de su propia media.

Finalmente, se puede probar la significación de una correlación múltiple (y de la correspondiente reducción proporcional del error) utilizando un procedimiento en el que la hipótesis nula establece que la correlación múltiple de la población es 0.

En este capítulo, nos referiremos a todo el procedimiento de regresión múltiple y correlación múltiple en su conjunto como "regresión/correlación múltiples". Es una costumbre ampliamente utilizada y simplifica la exposición.

INTRODUCCIÓN AL MODELO LINEAL GENERAL

Una forma de expresar el modelo lineal general es viéndolo como una relación matemática entre una variable dependiente y una o más variables de predicción. El principio básico establece que el valor de una persona en determinada variable dependiente (como por ejemplo el nivel de estrés) es la consecuencia de la suma de varias influencias:

1. Cierta influencia fija que será igual para todos los individuos, tales como la naturaleza del procedimiento de prueba o los impactos de la biología humana y la sociedad.

2. Influencias de otras variables que hemos medido en las que las distintas personas tienen registros diferentes, tales como cantidad de personas supervisadas, nivel de ruido y cantidad de decisiones por mes.

3. Otras influencias no medidas, que son las que producen el error.

La influencia 1 corresponde a la constante de regresión (a) en la ecuación de regresión múltiple. La influencia 2 corresponde a todos los pares de b y X, $(b_1)(X_1)$, $(b_2)(X_2)$, y así sucesivamente, en la ecuación de regresión múltiple. La influencia 3 se refiere a los errores de predicción. (Si existiera una correlación múltiple de 1,0 no existiría la influencia 3). Así, el modelo lineal general se puede expresar simbólicamente de la siguiente manera:

$$Y = a + (b_1)(X_1) + (b_2)(X_2) + (b_3)(X_3) + \ldots + e \qquad (16\text{-}2)$$

En la fórmula precedente, Y es el valor real de una persona en la variable dependiente. a es la influencia fija que se aplica a todos los individuos (influencia 1). b_1 es el grado de influencia de la primera variable de predicción (influencia 2); es decir, el coeficiente de regresión para puntuaciones originales, que después se multiplica por la puntuación original de la persona en la primera variable de predicción, X_1. b_2, b_3, y así sucesivamente, son las influencias de las variables de predicción 2, 3, y así respectivamente. e es el error, la suma de todas las otras influencias (influencia 3) en el valor de la persona en Y. Es decir, e es lo que queda después de haber tenido en cuenta todos los demás elementos de predicción.

La fórmula precedente, según se observa, es casi idéntica a la de la regresión múltiple, pero con dos excepciones. Primero, en lugar del valor Y predicho ($\hat{Y}$) ubicado a la izquierda, tenemos el valor real Y. Segundo, incluye el término de error (e), debido precisamente a que la fórmula busca el valor real de Y, y los valores a y b comúnmente no predicen en forma perfecta. El término de error (e) se agrega para justificar esa discrepancia.

Así, el modelo lineal general es la enunciación de las influencias que forman el valor de un individuo en una variable determinada. Se denomina **modelo lineal** porque si realizáramos un gráfico de la relación entre las variables dependiente y de predicción, la figura formada sería una línea recta. Es decir, la relación sería constante, en el sentido de que no es curvilínea. La influencia que actúa como tasa de cambio (el coeficiente de regresión) de cada variable de predicción siempre es la misma. En términos matemáticos, se dice que la ecuación es lineal porque no incluye términos elevados al cuadrado (o elevados a mayores potencias).[1]

Probablemente, el alumno también haya escuchado que varios procedimientos estadísticos utilizan un **modelo de cuadrados mínimos.** Se trata de un modelo en el que los valores a y b del modelo lineal general (o de una regla de predicción de regresión múltiple), para una variable dependiente en particular, se determinan de modo de crear la menor cantidad posible de error cuadrático, idea que ya hemos tratado extensamente.

MODELO LINEAL GENERAL Y REGRESIÓN/CORRELACIÓN MÚLTIPLES

El vínculo entre el modelo lineal general y la regresión/correlación múltiples es muy estrecho; son prácticamente lo mismo. Tradicionalmente, no se los ha equiparado porque se consideraba que el modelo lineal general estaba implícito en otras técnicas, tales como la correlación bivariada y el análisis de varianza, además de la regresión / correlación múltiples. Sin embargo, en los últimos años, los psicólogos han advertido (p. ej. Cohen & Cohen, 1983) que estas otras técnicas pueden derivar de la regresión / correlación múltiples al igual que del modelo lineal general.

REGRESIÓN Y CORRELACIÓN BIVARIADAS COMO CASOS ESPECIALES DE REGRESIÓN / CORRELACIÓN MÚLTIPLES

La regresión bivariada, es decir, la predicción de una variable dependiente a partir de una variable de predicción, es un caso especial de regresión múltiple, la predicción de una variable dependiente a partir de una cantidad cualquiera de variables de predicción. Asimismo, la correlación bivariada, la relación entre una variable de predicción y una variable dependiente, es un caso especial de correlación múltiple, la relación entre una cantidad cualquiera de variables de predicción y una variable dependiente.

LA PRUEBA *T* COMO CASO ESPECIAL DEL ANÁLISIS DE VARIANZA

La relación del modelo lineal general con la correlación y la regresión es bastante directa. La relación del modelo lineal general (o de la correlación y la regresión) con la prueba t y el análisis de varianza es menos directa. Sin embargo, antes de dedicarnos a esa relación, analicemos primero el vínculo entre la prueba t y el análisis de varianza.

[1] Existen métodos ingeniosos de introducir furtivamente términos elevados al cuadrado o a mayores potencias en los procedimientos del modelo lineal. Por ejemplo, podríamos crear una variable nueva, transformada, en la que cada valor estuviera elevado al cuadrado. Luego se podría utilizar esa variable transformada en una ecuación de modelo lineal como una variable original. Así, en la ecuación no aparecería en realidad ningún término elevado al cuadrado. Este pequeño truco resulta extraordinariamente valioso. Por ejemplo, ciertos textos sobre regresión múltiple (p. ej. Cohen & Cohen, 1983; Darlington, 1990) muestran la forma de utilizar ese tipo de procedimientos para trabajar con relaciones curvilíneas a través de métodos estadísticos diseñados para relaciones lineales.

Tanto la prueba t como el análisis de varianza son procedimientos para probar la diferencia entre medias de grupos. La prueba t se utiliza cuando existen sólo dos grupos.[2] El análisis de varianza con razón F, se utiliza generalmente sólo cuando existen más de dos grupos. Sin embargo, no existe motivo para no utilizar un análisis de varianza sólo con dos grupos. Cuando existen sólo dos grupos, la prueba t y el análisis de varianza producen conclusiones idénticas.

t y F son estrictamente idénticos sólo cuando se trabaja con dos grupos. Cuando existen más de dos grupos, no podemos realizar una prueba t ordinaria. Por eso decimos que la prueba t es un **caso especial** del análisis de varianza. La prueba t es matemáticamente idéntica al análisis de varianza en el *caso particular en el que existen sólo dos grupos (pronto analizaremos un ejemplo).*

Comprensión intuitiva de la relación entre los dos procedimientos

Un modo de percibir el vínculo entre los dos procedimientos es a través de la analogía con el coeficiente señal-ruido que presentamos en el capítulo 11 para explicar el análisis de varianza. La idea es que la razón F del análisis de varianza es una medida del grado en el cual la señal (análoga a la diferencia entre las medias de grupo) excede el ruido (análogo a la variación interna de cada uno de los grupos). La misma idea se aplica a la prueba t, que en realidad también determina el grado en el cual la señal (la diferencia entre las medias de los dos grupos) excede el ruido (el desvío estándar de la distribución de diferencias de medias, que también se basa en la variación interna de los grupos).

Paralelismos entre la lógica básica de los dos procedimientos

El análisis de varianza se basa en el cálculo de una razón F (que después se compara con el punto de corte tomado de una tabla basada en una distribución F). La razón F es la estimación de la varianza poblacional que se basa en la variación entre las medias de dos o más grupos, dividida por la estimación de la varianza poblacional basada en la variación dentro de cada uno de esos grupos. Es decir, la razón F es una fracción en la cual el numerador se basa en las diferencias entre los grupos comparando sus medias, y el denominador se basa en la variación dentro de cada uno de los grupos.

La prueba t se basa en el cálculo de un valor t (que después se compara con un punto de corte previamente definido, tomado de una tabla basada en una distribución t). El valor t es la diferencia entre las medias de los dos grupos dividida por el desvío estándar de la distribución de diferencias de medias. El desvío estándar de la distribución de diferencias de medias se calcula utilizando una estimación combinada de la varianza que se basa en el promedio de la varianza dentro de cada uno de los dos grupos. El valor t es una fracción en la que el numerador es la diferencia entre los grupos comparando sus medias, y el denominador se basa en la variación dentro de cada uno de los grupos.

En otras palabras, como lo indica la sección superior de la tabla 16-1, tanto una razón F como un valor t son fracciones en las cuales el numerador se basa en las diferencias entre las medias de los grupos y el denominador se basa en las varianzas dentro de los grupos.

[2] En este capítulo, nos concentramos en la prueba de hipótesis para medias independientes (y también en el análisis de varianza para diseños intersujetos). Sin embargo, las conclusiones finales son las mismas que con respecto a la prueba t para medias dependientes. Se trata de un caso especial de análisis de varianza de medidas repetidas. Además, tanto la prueba t para medias dependientes como el análisis de varianza de medidas repetidas, son casos especiales de regresión / correlación múltiples. De todos modos, el vínculo entre estos métodos y la correlación múltiple involucra algunos niveles extra de lógica que no analizamos aquí, para concentrarnos en las ideas principales del capítulo.

Cuadro 16-1. La época dorada de la estadística: cuatro muchachos en Londres.

En el último capítulo de su pequeño libro *Los pioneros de la estadística*, James Tankard (1984) trata el interesante hecho de que las cuatro técnicas estadísticas más comunes fueron creadas por cuatro ingleses nacidos dentro de un periodo de sesenta y ocho años, tres de los cuales trabajaban en las cercanías de Londres (y el cuarto, Gosset, a pesar de estar atascado en la fábrica de cerveza en Dublín, visitaba Londres para estudiar y se mantenía en contacto con todo lo que estaba sucediendo en esa ciudad). ¿Cuál era el motivo?

En primer lugar, Tankard sentía que la proximidad y comunicación de esos personajes fueron importantes para la creación de una "masa crítica" de mentes, que se suele relacionar con una época de oro para el descubrimiento y la creatividad. En segundo lugar, como sucede frecuentemente con los descubrimientos importantes, cada uno enfrentó complicados problemas prácticos o "anomalías" que los impulsaron hacia las soluciones que descubrieron. (Ninguno se propuso simplemente inventar un método estadístico). Galton (cuadro 3-1) estaba interesado en las características de padres e hijos; Pearson (cuadro 14-1) en la medición de la concordancia entre una serie de observaciones y una curva teórica. El problema de Gosset (cuadro 9-1) eran las pequeñas muestras ocasionadas por las condiciones económicas de la industria cervecera, y Fisher (cuadro 11-1) estaba estudiando los efectos del abono en las plantaciones de papas. (Tankard señala que la edad no fue un factor común. La edad en la que estos cuatro hombres realizaron su mayor contribución fluctúa entre los 31 y los 66 años).

Tankard también plantea tres importantes factores sociales específicos de esa "época dorada de la estadística". En primer lugar, hay que tener en cuenta el papel que desempeñaba la biometría, que intentaba probar matemáticamente la teoría de la evolución. La biometría ejerció influencia a través de la lectura que Galton hizo de Darwin y de la influencia de Galton sobre Pearson. En segundo lugar, este periodo presenció el comienzo de la contratación masiva, en el sector industrial y agrícola, de graduados universitarios con capacitación matemática avanzada. Y en tercer lugar, desde la época de Newton, la Universidad de Cambridge había sido una fuente especial y centralizada de matemáticos brillantes para Inglaterra. Podían diseminarse por toda la industria británica y, aún así, a través de su alma máter común permanecer en contacto tanto con los alumnos como entre sí, y también al tanto de los últimos descubrimientos.

Finalmente, Tankard dedica algunas palabras cálidas, casi poéticas, a la historia de esta ciencia en general y a su época dorada en particular:

> En realidad, es difícil comprender cómo puede rotularse la estadística como tediosa e inanimada. Después de escudriñar bajo la superficie de esta disciplina práctica y poderosa, comprendemos que, más de una vez, ha logrado producir fuertes pasiones y enérgicos debates entre las personas. Y siendo la estadística producto de la mente humana, sin duda continuará haciéndolo (p. 141).

Relación matemática entre los dos procedimientos

En los casos en los que hay sólo dos grupos, la fórmula para calcular el valor *t* es precisamente la raíz cuadrada de la fórmula para la razón *F*. A la mayoría de los alumnos no les interesará el origen preciso de esta relación, pero tiene una consecuencia importante. Si calculamos un valor *t*, será exactamente la raíz cuadrada del resultado que obtendríamos si calculáramos una razón *F* de la misma serie de observaciones. Por ejemplo, si calculáramos un *t* igual a 3 y después calculáramos un *F* con los mismos datos, el *F* sería igual a 9. Asimismo, veamos qué ocurre con los puntos de corte indicados en una tabla *t*: son exactamente la raíz cuadrada de los puntos de corte indicados en la columna de la tabla *F* que utilizamos cuando realizamos un análisis de varianza con dos grupos (es decir, en la parte de la tabla *F* en la que los grados de libertad del numerador son igual a 1).

Un aspecto particular de la equivalencia matemática de *t* y *F* ayudará a comprender el modo en que dos series de cálculos, aparentemente tan diferentes, encierran en realidad lo mismo. Una diferencia aparente entre los dos procedimientos es el modo en que los afecta el tamaño de la muestra. En el análisis de varianza, el tamaño de la muestra es parte del numerador. Tal como vimos en el capítulo 11, el numerador de la razón *F* es la estimación de la varianza poblacional que utiliza la diferencia entre las medias multiplicada por la cantidad de observaciones en cada grupo. Es decir, $S^2_{\text{Entre}} = (S^2_{M})(n)$. En la prueba *t*, el tamaño de la muestra es parte del denominador. Como vimos en el capítulo 10, el denominador de la prueba *t* utiliza la estimación combinada de la varianza poblacional dividida por la cantidad de observaciones de cada grupo.

Es decir, $S_{\text{diferencia}} = \sqrt{S^2_{\text{diferencia}}}$; $S^2_{\text{diferencia}} = S^2_{M1} + S^2_{M2}$; $S^2_{M1} = S^2_{\text{combinada}}/N_1$; $S^2_{M2} = S^2_{\text{combinada}}/N_2$).

Sin embargo, esta aparente contradicción se resuelve, porque multiplicar el numerador de una fracción por un número tiene exactamente el mismo efecto que dividir el denominador por ese mismo número. Por ejemplo, tomemos la fracción 3/8. Si multiplicamos el numerador por 2 nos da 6/8, ó 3/4; si dividimos el denominador de 3/8 por 2 también nos da 3/4.[3]

Tabla 16-1.
Algunos vínculos de la prueba *t* para medias independientes y el análisis de varianza.

Prueba *t*	**Análisis de varianza**
El numerador de *t* es la diferencia entre las medidas de dos grupos.	El numerador de *F* se basa, en parte, en la variación entre las medias de dos o más grupos.
El denominador de *t* se basa, en parte, en la combinación de las estimaciones de varianza poblacional calculadas a partir de cada grupo	El denominador de *F* se calcula combinando las estimaciones de varianza poblacional calculadas a partir de cada grupo.
El denominador de *t* implica dividir por la cantidad de registros.	El numerador de *F* involucra la multiplicación por la cantidad de observaciones. (Multiplicar un numerador por determinado número tiene el mismo efecto que dividir el denominador por ese mismo número).
Cuando se utilizan dos grupos, $t = \sqrt{F}$	Cuando se utilizan dos grupos, $F = t^2$
$gl = (N_1 - 1) + (N_2 - 1)$	$gl_{\text{Dentro}} = (n_1 - 1) + (n_2 - 1) + \ldots + (n_{\text{Último}} - 1)$

[3] Otras diferencias aparentes (tal como la supuesta diferencia entre el numerador de la razón *F*, que se basa en una estimación de varianza, y el numerador del punto *t*, que es una simple diferencia entre medias) presentan una unidad subyacente similar. Pero aquí no trataremos esos temas.

Tabla 16-2.
Cálculos de la prueba *t* y el análisis de varianza correspondientes a un experimento acerca de la efectividad de un nuevo programa de capacitación laboral (datos ficticios).

	Grupo experimental (Programa especial)			Grupo de control (Programa estándar)		
	X_1	$X_1 - M_1$	$(X_1 - M_1)^2$	X_2	$X_2 - M_2$	$(X_2 - M_2)^2$
	6	0	0	6	3	9
	4	–2	4	1	–2	4
	9	3	9	5	2	4
	7	1	1	3	0	0
	7	1	1	1	–2	4
	3	–3	9	1	–2	4
	6	0	0	4	1	1
Σ	42	0	24	21	0	26

$M_1 = 6$ $S_1^2 = 24/6 = 4$ $M_2 = 3$ $S_2^2 = 26/6 = 4{,}33$

$N_1 = 7$ $gl_1 = N_1 - 1 = 6$ $N_2 = 7$ $gl_2 = N_2 - 1 = 6$

Cálculos de la prueba *t* | **Cálculos del ANOVA**

Numerador

Diferencia media = 6,00 – 3,00 = 3,00

$gl_{\text{Entre}} = N_{\text{Grupos}} - 1 = 2 - 1 = 1$

$GM = (6 + 3)/2 = 9/2 = 4{,}5$

$\Sigma(M - GM)^2 = (6 - 4{,}5)^2 + (3 - 4{,}5)^2$
$= 1{,}5^2 + -1{,}5^2$
$= 2{,}25 + 2{,}25 = 4{,}5$

$$S^2_{\text{Entre}} \text{ ó } MS_{\text{Entre}} = \left(\frac{\Sigma(M - GM^2)}{gl_{\text{Entre}}}\right)(n) = \left(\frac{4{,}5}{1}\right)(7) = 31{,}5$$

Denominador

$$S^2_{\text{combinada}} = \left(\frac{gl_1}{gl_{\text{Total}}}\right)(S_1^2) + \left(\frac{gl_2}{gl_{\text{Total}}}\right)(S_2^2) = \left(\frac{6}{12}\right)(4) + \left(\frac{6}{12}\right)(4{,}33)$$
$$= (0{,}5)(4) + (0{,}5)(4{,}33) = 2{,}00 + 2{,}17 = 4{,}17$$

$$S^2_{\text{diferencia}} = S^2_{M_1} + S^2_{M_2} = (S^2_{\text{combinada}}/N_1) + (S^2_{\text{combinada}}/N_2)$$
$$= (4{,}17/7) + (4{,}17/7)$$
$$= 0{,}60 + 0{,}60 = 1{,}20$$

$$S_{\text{diferencia}} = \sqrt{S^2_{\text{diferencia}}} = \sqrt{1{,}20} = 1{,}10$$

$$S^2_{\text{dentro}} \text{ ó } CM_{\text{dentro}} = \frac{S_1^2 + S_2^2 + \ldots + S^2_{\text{Último}}}{N_{\text{Grupos}}} = \frac{4 + 4{,}33}{2}$$
$$= \frac{8{,}33}{2} = 4{,}17$$

Grados de Libertad

$gl_{\text{Total}} = gl_1 + gl_2 = 6 + 6 = 12$

$gl_{\text{dentro}} = gl_1 + gl_2 \ldots gl_{\text{Último}} = 6 + 6 = 12$

Corte

t necesario con $gl = 12$ a nivel 5%, dos colas= ±2,179

f necesario con $gl = 1{,}2$ a nivel 5%, = 4,75

Registro en la distribución comparativa

$t = (M_1 - M_2)/S_{\text{Diferencia}} = (6{,}00 - 3{,}00)/1{,}10 = 3{,}00/1{,}10 = 2{,}73$

$F = S^2_{\text{Entre}}/S^2_{\text{Dentro}}$ ó $CM_{\text{Entre}}/CM_{\text{Dentro}} = 31{,}5/4{,}17 = 7{,}55$

Tabla 16-2 (*cont.*).

Conclusiones	
Se rechaza la hipótesis nula. Se sostiene la hipótesis de investigación.	Se rechaza la hipótesis nula. Se sostiene la hipótesis de investigación.

Cálculos que ejemplifican la identidad de los dos procedimientos

La equivalencia se verá con mayor claridad a través de un ejemplo de los cálculos correspondientes. La tabla 16-2 muestra los cálculos de t y F para uno de los ejemplos de prueba t del capítulo 10. Se trata del experimento ficticio que prueba la efectividad de un nuevo programa de capacitación laboral para individuos que previamente no han sido capaces de mantener sus empleos. Es importante observar: a) La estimación combinada de la varianza poblacional de la prueba t ($S^2_{combinada} = 4{,}17$) es igual a la estimación intergrupal de la varianza poblacional del análisis de varianza ($S^2_{Dentro} = 4{,}17$), ambas calculadas como parte del denominador. b) Los grados de libertad de la distribución t ($gl = 12$) son exactamente iguales a los grados de libertad del denominador de la distribución F ($gl_{Dentro} = 12$). c) El punto de corte t para rechazar la hipótesis nula (2,179) es la raíz cuadrada del punto de corte F para rechazar la hipótesis nula ($\sqrt{4{,}75} = 2{,}179$). d) El estadístico t calculado con esta información (2,73) es la raíz cuadrada del F calculado ($\sqrt{7{,}55} = 2{,}75$, la diferencia se debe al redondeo). Y e) la conclusión es la misma: con ambos métodos rechazamos la hipótesis nula.

LA PRUEBA *T* COMO CASO ESPECIAL DE LA PRUEBA DE SIGNIFICACIÓN DEL COEFICIENTE DE CORRELACIÓN

La relación entre el coeficiente de correlación y la prueba t no es para nada obvia. Incluso, muchos investigadores psicológicos se han dado cuenta del vínculo recientemente. El coeficiente de correlación es el grado de relación entre dos variables; la prueba t trata sobre la significación de la diferencia entre dos medias poblacionales. ¿Cuál es la conexión posible? Una conexión es que ambos utilizan la distribución t para determinar la significación.

En el capítulo 3 todavía no habíamos analizado la lógica de la prueba de hipótesis, por lo tanto, sólo podíamos exponer la significación de un coeficiente de correlación en términos muy generales. Con todo lo que ya hemos aprendido, ahora podemos explicarla con más precisión. El procedimiento sigue los cinco pasos estándar de prueba de hipótesis. Sus características principales son: a) la hipótesis nula establece que la población tiene una correlación igual a 0; b) la distribución comparativa es una distribución t con tantos grados de libertad como la cantidad de participantes menos 2, y c) el valor en la distribución comparativa es un valor t calculado a partir del coeficiente de correlación utilizando la fórmula: $t = r\sqrt{N-2}/\sqrt{1-r^2}$. (Para mayores detalles, incluso un ejemplo y el análisis de cuestiones relacionadas con el tamaño de efecto y la potencia, *véase* el apéndice II del capítulo 3). Es importante señalar que la clave de todo el proceso es convertir el coeficiente de correlación en un valor t.

Sin embargo, conocer este procedimiento no nos permite discernir con claridad por qué se puede convertir al coeficiente de correlación en un valor t con el propósito de realizar una prueba de hipótesis. Tampoco nos explica la conexión entre el t basado en el coeficiente de correlación y la prueba t aplicada para probar la diferencia entre las medias de dos grupos. A estos temas nos dedicaremos ahora.

Diferencias grupales expresadas como relaciones entre variables

Generalmente pensamos en el coeficiente de correlación como la relación entre una variable de predicción (o independiente) y una variable dependiente. Probar la significación de un coeficiente de correlación implica preguntarse si podemos rechazar la hipótesis nula que establece que en la población no existe relación entre las dos variables (que en la población, $r = 0$).

La prueba t para medias independientes analiza la diferencia entre dos medias poblacionales sobre la base de las medias de dos muestras. Las muestras se miden según una variable dependiente. Lo que diferencia a los dos grupos es la variable independiente o de predicción. En nuestro ejemplo de la sección anterior, la variable independiente era la que indicaba si los participantes recibían el nuevo programa de capacitación laboral o el programa ordinario. La hipótesis nula puesta a prueba establece que el grupo al que pertenece el participante (la variable de predicción) no tiene ningún efecto sobre la variable dependiente. La prueba t está analizando si en la población en general existe alguna relación entre la variable de predicción y la dependiente. Si volvemos a analizar la oración anterior, veremos que se ha reconstruido la prueba t convirtiéndo-la en una relación entre la variable de predicción y una variable dependiente.

En otras palabras, un coeficiente de correlación significativo indica que la variable de predicción y la dependiente están relacionadas. Una prueba t de medias independientes, que resulta significativa, indica que la variable de predicción y la dependiente están relacionadas. Ambas indican lo mismo.

Variables numéricas de predicción en comparación con variables nominales de predicción dicótomicas

A esta altura el alumno podría objetar: "La variable de predicción en un coeficiente de correlación es una variable numérica, tal como la cantidad de personas supervisadas o el promedio de calificaciones en el colegio secundario. La variable de predicción en una prueba t para medias independientes es una variable con exactamente dos valores, las dos categorías, tales como un grupo experimental en comparación con un grupo de control"

Sí, es verdad. Esa es precisamente la diferencia entre los casos en los que utilizamos un coeficiente de correlación y aquellos en los que utilizamos una prueba t para medias independientes.

Generalmente, tanto para el coeficiente de correlación como para la prueba t para medias independientes, la variable dependiente puede tener un rango de valores numéricos. Sin embargo, con respecto a la variable de predicción existe una diferencia. En el caso del coeficiente de correlación, la variable de predicción, al igual que la variable dependiente, también es típicamente numérica. Por ejemplo, una correlación entre cantidad supervisada y nivel de estrés es una relación entre variables numéricas. En una prueba t para medias independientes la situación es diferente; en ese caso, la variable de predicción tiene exactamente dos valores, los que de ningún modo son números. La variable de predicción en una prueba t es una variable nominal con sólo dos valores (dicotómicas). Tiene dos categorías distintas, como por ejemplo, pertenecer al grupo experimental o al grupo de control.

Cerrando la brecha entre las variables numéricas y las variables nominales dicotómicas

¿Cómo podríamos cerrar esa brecha? Supongamos que otorgamos arbitrariamente dos números a la variable nominal con dos categorías. Por ejemplo, llamemos 1 al grupo experimental y 2 al grupo de control. (Utilizar otros dos números cualesquiera, finalmente produce exactamente el mismo resultado al convertir todo en puntuaciones Z para calcular el coeficiente de correlación. Los dos números que utilicemos y, específicamente, a qué categoría apliquemos el número más alto, determinará simplemente el signo positivo o negativo del resultado final).

Una vez que hemos convertido de este modo una variable de predicción nominal, con dos categorías para una prueba *t* de medias independientes, en una variable numérica (con sólo dos valores, debemos admitir), podemos proceder a calcular el coeficiente de correlación y, finalmente, determinar su significación.

Ejemplo de la equivalencia de cálculo entre la prueba *t* y la prueba de significación del coeficiente de correlación

La tabla 16-3 indica los cálculos del coeficiente de correlación y su significación, con los valores del mismo ejemplo de prueba *t* que utilizamos anteriormente. (Para que la tabla fuera razonable-

Tabla 16-3.
Cálculo del coeficiente de correlación y de una prueba de hipótesis sobre el coeficiente de correlación con los datos de la tabla 10-3 (y tabla 16-2), en el que se convierte la variable de predicción (independiente) en una variable numérica con los valores 1 (para el grupo experimental) ó 2 (para el grupo de control).

Variación de predicción (experimental versus control)		**Variable dependiente (calificación del empleador)**		**Producto cruzado**
Ordinario	Z_X	*Ordinario*	Z_Y	Z_XZ_Y
1	–1	6	0,62	–0,62
1	–1	4	–0,21	0,21
1	–1	9	1,87	–1,87
1	–1	7	1,04	–1,04
1	–1	7	1,04	–1,04
1	–1	3	–0,62	0,62
1	–1	6	0,62	–0,62
2	1	6	0,62	0,62
2	1	1	–1,45	–1,45
2	1	5	0,21	0,21
2	1	3	–0,62	–0,62
2	1	1	–1,45	–1,45
2	1	1	–1,45	–1,45
2	1	4	–0,21	–0,21
Σ 21	0	63	0	–8,71
M = 1,5	0	4,5	0	*r* = –0,62
(*SD* = 0,5)		(*SD* = 2,41)		

$gl = N - 2 = 14 - 2 = 12.$

Punto de corte *t* con $gl = 12$ a nivel 5%, dos colas = ±2,179.

$t = r\sqrt{N-2}/\sqrt{1-r^2} = -0{,}62\sqrt{14-2}/\sqrt{1-(-0{,}62)^2} = -0{,}62\sqrt{12}/\sqrt{1-0{,}38} = -0{,}62(3{,}46)/\sqrt{0{,}62} = -2{,}15/0{,}79 = -2{,}72$

Conclusión: Se rechaza la hipótesis nula; se sostiene la hipótesis de investigación.

mente simple, hemos excluido los cálculos del desvío estándar de cada variable, que se utilizan para calcular las puntuaciones Z). Es importante señalar que, en esta estructura de correlación, a cada individuo le corresponden dos observaciones: a) un 1 ó un 2, según la persona se encuentre en el grupo experimental (el grupo que recibe el nuevo programa de capacitación laboral) o en el grupo de control (el grupo que recibe el programa estándar), y b) un valor en la variable dependiente, que es la calificación de desempeño laboral determinada por el empleador un mes más tarde. La correlación resultante es –0,62. Aplicando la fórmula para convertir una correlación en una puntuación *t*, obtenemos un *t* de –2,72. Es el mismo *t* que habíamos calculado anteriormente (2,73) utilizando los procedimientos de la prueba *t* (tablas 10-3 y 16-2). La pequeña diferencia se debe al redondeo. La diferencia de signos está relacionada con el hecho de que a un grupo se le adjudica arbitrariamente el número 1 y al otro el número 2. Los grados de libertad, y por ende el punto de corte *t* necesario para alcanzar la significación y llegar a una conclusión, también son iguales a los utilizados cuando calculamos con los mismos datos una prueba *t* para medias independientes.

Tal como lo ilustra el presente ejemplo, la prueba de significación del coeficiente de correlación da el mismo resultado que la prueba *t* común. Sin embargo, decimos que la prueba *t* es un caso especial del coeficiente de correlación, porque la prueba *t* es sólo una instancia particular del coeficiente de correlación, es decir, es la situación en la que la variable de predicción tiene sólo dos valores.

Interpretación gráfica de la relación de la prueba *t* con el coeficiente de correlación

Analizando la situación gráficamente, podemos ver con mayor agudeza la relación entre *t* y *r*. Es posible realizar un gráfico con los datos de una prueba *t* para medias independientes a través de un diagrama de dispersión, como lo haríamos para un coeficiente de correlación con una variable de predicción de sólo dos valores. De hecho, los gráficos realizados con la misma información son iguales. Analicemos uno. La figura 16-2 muestra el diagrama de dispersión con la correspondiente recta de regresión que representa los datos del estudio acerca de la capacitación laboral. La variable de predicción tiene sólo dos valores; por lo tanto, en el diagrama de dispersión todos los puntos se alinean sobre esos dos valores. Es importante resaltar que la recta de regresión pasa a

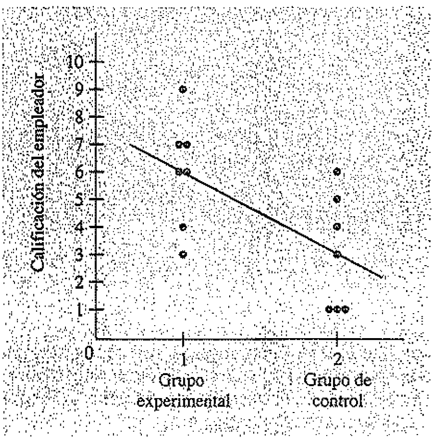


Figura 16-2.
Diagrama de dispersión y recta de regresión del ejemplo acerca de la capacidad laboral, originalmente analizado con una prueba *t* para medias independientes, con un valor de 1 para el grupo experimental y 2 para el grupo control.

través del medio de cada serie de puntos. En realidad, si al realizar un diagrama de dispersión con los resultados de una prueba t, calculáramos la recta de regresión, esa recta siempre caería exactamente en la media de cada serie de puntos. Es decir, la recta de regresión pasa por la ubicación que representa la media de cada grupo, ya que en cada serie de observaciones, el mejor número de predicción es siempre la media (en el sentido de producir el mínimo error cuadrático).

Ahora analicemos algunos patrones posibles en este tipo de diagrama de dispersión. La figura 16-3a representa un caso en el que las dos medias son casi iguales. En ese caso, la pendiente de la recta de regresión es prácticamente 0; la correlación es baja y no es significativa. De hecho, con los datos del ejemplo, la correlación es 0,10. Utilizando la fórmula de la prueba t para el coeficiente de correlación, con 20 participantes, el resultado es un t de 0,43:
$t = r\sqrt{N-2} / \sqrt{1-r^2} = 0{,}1\sqrt{20-2} / \sqrt{1-0{,}12} = 0{,}43.$

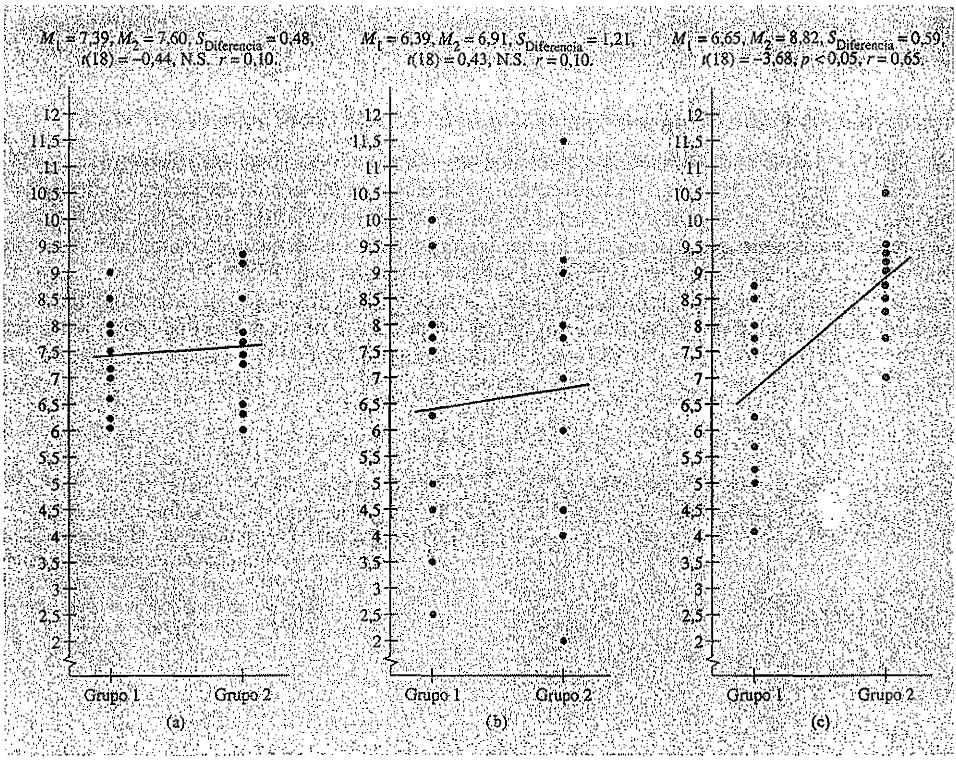


Figura 16-3.
Tres posibles diagramas de dispersión de datos analizados con una prueba t para medias independientes, en los que las medias de los dos grupos son (a) casi iguales; (b) diferentes pero con datos que están muy dispersos (gran varianza combinada o gran desvío estándar de la distribución de la diferencia de medias muestrales), y (c) muy diferentes, con datos que no están ampliamente dispersos.

Del mismo modo, pensando en términos de una prueba t para medias independientes, habiendo tan poca diferencia entre las medias de los dos grupos, la prueba t tampoco será significativa. Los datos del ejemplo presentan una diferencia de media de 7,39 – 7,60 = 0,21. Con un desvío estándar de la distribución de diferencias entre medias de 0,48, el t es de –0,44: $t = (M_1 - M_2)/S_{\text{Diferencia}}$ = (7,39 – 7,60)/0,48 = –0,44. Con diferencia de redondeo (e ignorando el signo), es el mismo resultado que obtenemos utilizando el método de la correlación.

La figura 16-3b representa un caso en el que las medias de los dos grupos son algo diferentes pero en donde los puntos de cada grupo están aún más dispersos. En ese caso, nuevamente la recta de regresión es un elemento de predicción muy pobre. Una vez más, el coeficiente de correlación, aunque no es igual a 0, de todos modos sería bastante bajo y no significativo. De hecho, con los datos del ejemplo, r = 0,10, el cuál no es estadísticamente significativo. En la prueba t para medias independientes realizada con estos mismos datos, el efecto de la dispersión de los puntos es una mayor varianza poblacional estimada para cada grupo. A la vez, lo anterior causa una estimación de varianza combinada considerable y un gran desvío estándar de la distribución de diferencias entre medias. Dado que en una prueba t se divide la diferencia de medias por el desvío estándar de la distribución de diferencias entre medias, cuanto más grande es el desvío estándar, menor será el t que resulte. Los datos del ejemplo arrojan una diferencia de medias de 0,52 y un desvío estándar de la distribución de diferencia de medias de 1,21. El resultado es un t de 0,43 que, claramente, no es significativo.

Por el contrario, la figura 16-3c representa un caso en el que existe una gran diferencia entre las medias con una variación relativamente pequeña entre los puntos que rodean a cada media. Como resultado, la línea de regresión es muy útil como elemento de predicción, dando un alto coeficiente de correlación. (Aplicando los datos del ejemplo, r = 0,65 y t = 3,63, según se calcula a partir de r para probar su significación). Asimismo, la gran diferencia de media y la pequeña varianza dentro de cada grupo contribuyen a un gran t cuando se calcula utilizando una prueba t para medias independientes. En este ejemplo, la diferencia media es –2,17 y el desvío estándar de la distribución de diferencias entre medias es 0,59; por lo tanto, t es –3,68 (la diferencia con el t calculado utilizando r se debe al redondeo).

El principio representado gráficamente por las figuras que acabamos de analizar es que la prueba t para medias independientes y la prueba de significación del coeficiente de correlación dan los mismos resultados, porque ambas son mayores cuando la diferencia entre las dos medias es grande y la variación entre las observaciones de cada grupo es pequeña.

EL ANÁLISIS DE VARIANZA COMO CASO ESPECIAL DE LA PRUEBA DE SIGNIFICACIÓN DEL COEFICIENTE DE CORRELACIÓN MÚLTIPLE

La relación entre el análisis de varianza y la correlación múltiple es paralela a la relación que acabamos de analizar entre la prueba t para medias independientes y el coeficiente de correlación (bivariado) ordinario. En ambas relaciones, uno de los dos estadísticos parece referirse a las diferencias entre medias y el otro a las asociaciones entre variables. La resolución de esta diferencia aparente es la misma. El análisis de varianza analiza si existe una diferencia, en la variable dependiente, entre las medias de los grupos que representan diferentes niveles de una variable de predicción (o independiente). El método de la correlación encara la situación como una relación entre la variable dependiente y los diferentes niveles de la variable de predicción. Por ejemplo, en el estudio de Hazan y Shaver (1987) que analizamos en el capítulo 11, acerca del estilo de relación y los celos, los investigadores descubrieron que el análisis de varianza mostraba una diferencia signifi-

Tabla 16-4.
Cálculo de la reducción proporcional del error con observaciones originales y análisis de varianza, método del modelo estructural, con los datos de la tabla 10-3.

REDUCCIÓN PROPORCIONAL DEL ERROR

Variable de predicción (experimental versus control)	Variable dependiente (calificación del empleador)			
Código	*Calificación*	*Predicho*	*Diferencia*	*Diferencia cuadrática*
1	6	6	0	0
1	4	6	−2	4
1	9	6	3	9
1	7	6	1	1
1	7	6	1	1
1	3	6	−3	9
1	6	6	0	0
2	6	3	3	9
2	1	3	−2	4
2	5	3	2	4
2	3	3	0	0
2	1	3	−2	4
2	1	3	−2	4
2	4	3	1	1
				$\Sigma = SS_{\text{Error}} = 50$

Suma de cuadrados utilizando la media general
como regla de predicción (no se muestra el cálculo): $SS_{\text{Total}} = 81{,}5$

$$\text{Reducción proporcional del error cuadrático} = \frac{SS_{\text{Total}} - SS_{\text{Error}}}{SS_{\text{Total}}} = \frac{81{,}5 - 50}{81{,}5} = \frac{31{,}5}{81{,}5} = 0{,}39$$

$r^2 = 0{,}39; r = \sqrt{r^2} = \sqrt{0{,}39} = \pm 0{,}62.$

CÁLCULO DEL MODELO ESTRUCTURAL DE UN ANÁLISIS DE VARIANZA DE UN CRITERIO

$GM = 4{,}5$

	Grupo experimental (programa especial)							Grupo de control (programa estándar)					
X_1	$X - GM$		$X - M$		$M - GM$		X	$X - GM$		$X - M$		$M - GM$	
	Desv	*Desv²*	*Desv*	*Desv²*	*Desv*	*Desv²*		*Desv*	*Desv²*	*Desv*	*Desv²*	*Desv*	*Desv²*
6	1,5	2,25	0	0	1,5	2,25	6	1,5	2,25	3	9	−1,5	2,25
4	−0,5	0,25	−2	4	1,5	2,25	1	−3,5	12,25	−2	4	−1,5	2,25
9	4,5	20,25	3	9	1,5	2,25	5	0,5	0,25	2	4	−1,5	2,25
7	2,5	6,25	1	1	1,5	2,25	3	−1,5	2,25	0	0	−1,5	2,25
7	2,5	6,25	1	1	1,5	2,25	1	−3,5	12,25	−2	4	−1,5	2,25
3	−1,5	2,25	−3	9	1,5	2,25	1	−3,5	12,25	−2	4	−1,5	2,25
6	1,5	2,25	0	0	1,5	2,25	4	−0,5	,25	1	1	−1,5	2,25
Σ:		39,75		24		15,75	Σ:		41,75		26		15,75

Nota: Desv = Desvío; Desv² = Desvío cuadrático

Tabla 16-4 (cont.).

Suma de desvíos cuadráticos:

$\Sigma(X - GM)^2$ ó $SS_{\text{Total}} = 39{,}75 + 41{,}75 = 81{,}5$

$\Sigma(X - M)^2$ ó $SS_{\text{dentro}} = 24 + 26 = 50$

$\Sigma(M - GM)^2$ ó $SS_{\text{entre}} = 15{,}75 + 15{,}75 = 31{,}5$

Control ($SC_{\text{Total}} = SC_{\text{entre}} + SC_{\text{entre}}$): $81{,}5 = 50 + 31{,}5$

Grados de libertad:

$gl_{\text{Total}} = N - 1 = 14 - 1 = 13$

$gl_{\text{dentro}} = gl_1 + gl_2 + \ldots + gl_{\text{Último}} = 6 + 6 = 12$

$gl_{\text{entre}} = N_{\text{grupos}} - 1 = 2 - 1 = 1$

Control ($gl_{\text{Total}} = gl_{\text{Dentro}} + gl_{\text{Entre}}$): $13 = 12 + 1$

Estimaciones de varianza poblacional:

S^2_{Total} ó $CM_{\text{Total}} = SC_{\text{Total}}/gl_{\text{Total}} = 81{,}5/13 = 6{,}27$

S^2_{dentro} ó $CM_{\text{dentro}} = SC_{\text{dentro}}/gl_{\text{dentro}} = 50/12 = 4{,}17$

S^2_{entre} ó $CM_{\text{entre}} = SC_{\text{entre}}/gl_{\text{entre}} = 31{,}5/1 = 31{,}5$

Razón ó F: $F = S^2_{\text{entre}}/S^2_{\text{dentro}}$ ó $CM_{\text{entre}}/CM_{\text{dentro}} = 31{,}5/4{,}17 = 7{,}55$

$R^2 = \text{eta}^2 = SC_{\text{entre}}/SC_{\text{Total}} = 31{,}5/81{,}5 = 0{,}39$

cativa entre los tres estilos de relación (la variable independiente o de predicción) con respecto a los celos (la variable dependiente). Por el contrario, un enfoque correlacional describiría este resultado como una asociación significativa entre la variable del estilo de relación y la variable de los celos.

Análisis de varianza para dos grupos como caso especial de significación de una correlación bivariada

El vínculo entre el análisis de varianza y la correlación es más fácil de captar si interpretamos al coeficiente de correlación como la raíz cuadrada de la reducción proporcional del error con observaciones originales (*véase* capítulo 4), y al análisis de varianza utilizando el método del modelo estructural (capítulo 12). La parte superior de la tabla 16-4 indica los datos de la correlación del ejemplo acerca del experimento realizado con la capacitación laboral. Sin embargo, esta vez aparecen códigos asignados a los grupos experimental y control, a las observaciones originales, a los valores predichos y a los errores cuadráticos, al igual que a los cálculos para la reducción proporcional del error. La parte inferior de la tabla 16-4 indica los cálculos del análisis de varianza con los mismos datos, utilizando el método del modelo estructural.

Existen varios vínculos claros. Primero, la suma de los errores cuadráticos, calculada en la correlación cuando se utiliza la regla de predicción bivariable ($SC_{\text{Error}} = 50$), es igual a la suma de desvíos cuadráticos intragrupales (SC_{dentro}) correspondientes al análisis de varianza. ¿Por qué son iguales? El análisis de correlación está calculando el error como la diferencia con respecto al valor predicho, y el valor predicho es la media de cada grupo. Es decir, en el análisis de correlación la suma de los errores cuadráticos es el resultado de elevar al cuadrado y sumar la diferencia entre cada valor y la media de su grupo (que es la predicción para cada registro en su grupo). El análisis de varianza está calculando la suma de los errores cuadráticos intragrupales exactamente del mismo modo, la suma de los desvíos cuadráticos de cada observación con respecto a la media de su grupo.

Segundo, la suma de los errores cuadráticos en el análisis de correlación, cuando para predecir se utiliza la media general de la variable dependiente ($SC_{\text{Total}} = 81{,}5$), es igual a SC_{Total} en el análisis de varianza (también 81,5). Son iguales porque el análisis de correlación está calculando este error como el desvío cuadrático de cada observación con respecto a la media general de todas las observaciones de la variable dependiente, y el análisis de varianza está calculando la suma de los desvíos cuadráticos de cada observación con respecto a la gran media, que es la media general de todas las observaciones de la variable dependiente.

Tercero, la reducción del error cuadrático –la suma de cuadrados utilizando la media para predecir (81,5) menos la suma de cuadrados del error utilizando la regla de predicción bivariada (50)– es igual a 31,5. Este resultado coincide con la suma de cuadrados intergrupales (SC_{entre}, que es igual a 31,5) en el análisis de varianza. La reducción de error en el análisis correlacional es equivalente a lo que agrega la regla de predicción con respecto a conocer sólo la media. En este caso, la recta de predicción estima la media de cada grupo; por lo tanto, la reducción del error cuadrático de cada observación es la diferencia cuadrática entre la media del grupo de esa observación y la media general. SC_{entre}, en el análisis de varianza, se calcula sumando, por cada participante, las diferencias cuadráticas entre la media del grupo del participante y la gran media.

Finalmente, la reducción proporcional del error (r^2, también denominada proporción de varianza explicada), en el análisis de correlación, es exactamente igual a la proporción de varianza explicada (R^2 o eta^2), una de las medidas del tamaño del efecto que estudiamos en el análisis de varianza (ambas son igual a 0,39).

También, ambas nos indican la proporción de la variación total en la variable dependiente, que se explica a través de su relación con la variable independiente. A esta altura no debería sorprendernos que estos números sean iguales; ya hemos visto que los términos que forman tanto el numerador como el denominador, son los mismos en las fórmulas de r^2 y R^2.

Según lo observado, los vínculos entre correlación y análisis de varianza son bastantes profundos. De hecho, algunos investigadores calculan la significación de un coeficiente de correlación insertando las distintas sumas de cuadrados que éste produce en una tabla de análisis de varianza y calculando F. El resultado es idéntico al producido por cualquier otro método de cálculo de la significación del coeficiente de correlación. Si calculamos el t correspondiente a la correlación, el resultado es la raíz cuadrada del F que obtendríamos utilizando ese procedimiento.

Análisis de varianza para más de dos grupos como caso especial de correlación múltiple

Recordemos la táctica que empleamos al analizar la prueba t para medias independientes, como caso especial de prueba de significación del coeficiente de correlación. En ese caso pudimos calcular un coeficiente de correlación con los datos de la prueba t, convirtiendo arbitrariamente las dos categorías de la variable nominal de predicción en dos números distintos cualesquiera (en el ejemplo, utilizamos 1 para el grupo experimental y 2 para el grupo de control). El problema es más complejo cuando la variable de predicción incluye más de dos categorías, como sucede en un análisis de varianza para más de dos grupos.

Tuvimos éxito al asignar arbitrariamente dos números cualesquiera a las dos categorías, porque, en ese caso, no tienen importancia los números en particular –sólo deben ser diferentes. Cuando calculamos una correlación convertimos las observaciones de cada variable en puntuaciones Z, y el proceso tiene en cuenta el grado de diferencia entre los dos números. Pero asignar cualquier número arbitrariamente, cuando existen tres o más grupos, no funciona. Cualesquiera sean los tres números que elijamos, implican alguna relación particular entre los grupos, y no todas las relaciones serán iguales.

En el capítulo 11, analizamos un ejemplo en el que el investigador comparaba los distintos grados de culpabilidad de un acusado indicados por participantes, bajo tres condiciones: los que creían que el acusado tenía antecedentes delictivos; los que creían que no tenía antecedentes, y aquellos a los que no se les había hecho mención de los antecedentes. Supongamos que asignáramos arbitrariamente un 1 al primer grupo, un 2 al segundo y un 3 al tercero. Esto implicaría que consideramos que esos tres niveles son valores igualmente distanciados de una variable numérica que representa el conocimiento de los antecedentes delictivos. Convertir los valores 1, 2 y 3 en puntuaciones Z no sería una solución, ya que, de todos modos, continuarían dispersos en forma pareja y en el mismo orden. En este ejemplo en particular podríamos pretender considerar que los tres grupos están ordenados, comenzando por el de antecedentes delictivos y finalizando con el que fue informado de la ausencia de antecedentes, quedando el grupo que no recibió información en medio de los otros dos. Sin embargo, aun así no quedaría claro que en esta dimensión los grupos estén separados en forma pareja.

Generalizando, cuando trabajamos con tres grupos, no tenemos una base preestablecida para ubicar a los grupos en un orden determinado, mucho menos para decidir cómo deberían dispersarse. Por ejemplo, en un estudio que compara actitudes de cuatro nacionalidades centroamericanas diferentes, la nacionalidad es la variable de predicción; sin embargo, las cuatro nacionalidades no pueden convertirse en cuatro valores de una sola variable numérica.

De todos modos, existe una solución inteligente para este problema. En lugar de intentar transformar una variable nominal, con más de dos categorías, en una sola variable numérica, podemos convertirla en varias variables numéricas con dos niveles cada una.

El procedimiento es el siguiente: supongamos que la variable de predicción tiene cuatro categorías, por ejemplo, cuatro nacionalidades centroamericanas: costarricense, guatemalteca, nicaragüense y salvadoreña. Una variable de predicción podría ser el hecho de que el participante fuera costarricense: 1 si lo es y 0 si no lo es. Una segunda variable de predicción sería el hecho de que el participante fuera guatemalteco: 1 ó 0. Una tercera variable de predicción sería el hecho de que el participante fuera nicaragüense: 1 ó 0. Podríamos tener incluso una cuarta variable que establezca que el participante es salvadoreño; sin embargo, sucede que, en este ejemplo, si un participante tiene valor 0 en las primeras tres variables, por eliminación debe ser salvadoreño (porque existen sólo cuatro posibilidades). En general, identificar todos los niveles de una variable nominal requiere una variable menos que la cantidad de niveles de la variable nominal.

Volviendo al ejemplo, el resultado del procedimiento descripto es que la nacionalidad de cualquier participante se describe a través de los valores en tres variables numéricas, cada una con las posibilidades de ser 1 ó 0. Por ejemplo, un participante costarricense tendría un 1 en la opción costarricense y ceros en las opciones guatemalteco y nicaragüense. Cada participante guatemalteco tendría un 1 en la opción guatemalteca y ceros en las opciones costarricense y nicaragüense. Cada

Tabla 16-5.
Ejemplo de codificación nominal para la nacionalidad de diez participantes en un estudio ficticio de participantes de cuatro nacionalidades centroamericanas.

Participante	Nacionalidad	Variable 1 Costaricense o no	Variable 2 Guatemalteca o no	Variable 3 Nicaragüense o no
1	Guatemalteca	0	1	0
2	Nicaragüense	0	0	1
3	Salvadoreña	0	0	0
4	Nicaragüense	0	0	1
5	Costarricense	1	0	0
6	Costarricense	1	0	0
7	Salvadoreña	0	0	0
8	Nicaragüense	0	0	1
9	Costaricense	1	0	0
10	Guatemalteca	0	1	0

Tabla 16-6.
Ejemplo de codificación nominal para la condición experimental de quince participantes del ejemplo de antecedentes delictivos (datos ficticios).

		Variable de predicción o independiente		**Variable dependiente**
Participante	*Condición experimental*	*Variable 1: antecedentes delictivos o no*	*Variable 2: sin antecedentes o no*	*Calificación del participante con respecto a la culpabilidad del acusado*
1	Antecedente delictivo	1	0	10
2	Antecedente delictivo	1	0	7
3	Antecedente delictivo	1	0	5
4	Antecedente delictivo	1	0	10
5	Antecedente delictivo	1	0	8
6	Sin antecedentes	0	1	5
7	Sin antecedentes	0	1	1
8	Sin antecedentes	0	1	3
9	Sin antecedentes	0	1	7
10	Sin antecedentes	0	1	4
11	Sin información	0	0	4
12	Sin información	0	0	6
13	Sin información	0	0	9
14	Sin información	0	0	3
15	Sin información	0	0	3

participante nicaragüense tendría ceros en las opciones costarricense y guatemalteca. Cada participante salvadoreño tendría ceros en las tres variables. (A propósito, puede utilizarse cualquier par de números para cada variable nominal de dos valores; utilizamos 1 y 0 sólo por conveniencia). La tabla 16-5 muestra el funcionamiento de esta codificación aplicada a 10 participantes ficticios.

Todo el procedimiento descripto se denomina **codificación nominal**. (Convertir en 1 y 2 a los niveles del ejemplo de la prueba *t*, para calcular un coeficiente de correlación, también fue un caso de codificación nominal para una variable nominal de dos categorías). En el ejemplo que estamos analizando ahora, el resultado de la codificación nominal es que la variable de predicción, en lugar de ser una variable nominal con cuatro categorías, ahora se ha convirtido en tres variables numéricas pero con sólo dos valores cada una. Crear una serie de variables numéricas con dos valores, tal como acabamos de describir, evita el inconveniente de crear una jerarquización falsa de los cuatro niveles.

La tabla 16-6 muestra otro ejemplo de codificación nominal, esta vez aplicado a los participantes del ejemplo relacionado con los antecedentes delictivos. El resultado es que la variable de predicción, en lugar de ser una variable nominal con tres categorías, ahora se transformó en dos variables numéricas (con sólo dos valores cada una, 0 ó 1). Generalizando, en un análisis de varianza se puede codificar toda variable independiente nominal para convertirla en una serie de variables numéricas de dos valores. La serie estará formada exactamente por una variable menos que la cantidad de niveles que tenía la variable nominal. (No es coincidencia que resulte el mismo número que los grados de libertad de la estimación intergrupal de varianza poblacional).

Esa capacidad para codificar una variable nominal independiente, y convertirla en una serie de variables numéricas de dos valores en el análisis de varianza, es una transición importante que hace posible la realización de un análisis de correlación múltiple. Tomemos nuevamente el ejemplo de los antecedentes delictivos. Habiendo realizado la codificación nominal, ahora podemos calcular la correlación múltiple de las dos variables numéricas de predicción junto con la variable dependiente, el nivel de culpabilidad. El resultado final (en términos de nivel de significación y R^2) será idéntico al del análisis de varianza.

El procedimiento de codificación nominal que hemos descripto aquí implica la conversión de una variable nominal de predicción de un análisis de varianza, en distintas variables numéricas de dos niveles para una correlación múltiple.

Este procedimiento es extremadamente flexible y puede extenderse a los casos más complejos del análisis factorial de varianza. En verdad, lo importante no es que podamos realizar una codificación nominal; en la mayoría de los casos, una computadora lo hará por nosotros. Lo realmente importante es comprender el principio que hace posible la conversión de un problema de análisis de varianza en un problema de regresión múltiple. (Si el alumno está interesado en el tema de la codificación nominal, *véase* Cohen & Cohen, 1983, capítulo 5. Incluye una descripción detallada y de fácil lectura).

ELECCIÓN DE PRUEBAS ESTADÍSTICAS

Hemos visto que los cuatro procedimientos estadísticos principales que hemos aprendido en este libro se pueden considerar casos especiales de regresión / correlación múltiples. El alumno se preguntará por qué no aprendemos sólo una técnica, regresión / correlación múltiple, y resolvemos todos los casos utilizando esa misma técnica. Podríamos –y si lo hiciéramos obtendríamos resultados completamente correctos en todos los casos. De hecho, tal como lo indicamos al comienzo, si tuviéramos que aprender sólo un procedimiento para analizar datos estadísticos por computadora, deberíamos aprender la regresión / correlación múltiples.

Entonces, ¿por qué alguien utilizaría, digamos, una prueba *t* en lugar de un análisis de varianza? Simplemente porque es un procedimiento tradicional y ampliamente comprendido. En la actualidad, cuando se comparan dos grupos, la mayoría de los investigadores esperan encontrar una prueba *t*. (De todos modos, la situación está cambiando rápidamente a medida que los investigadores se vuelven cada vez más sofisticados). Resulta extraño, y de algún modo ampuloso, ver un análisis de varianza donde funcionaría una prueba *t*, aunque, de hecho, considerarlo ampuloso es un resabio de aquellos tiempos en los que se realizaban los cálculos a mano, y un análisis de varianza era más difícil de realizar que una prueba *t*.

Utilizar un coeficiente de correlación (y la prueba de significación correspondiente) en un caso con dos grupos, en lugar de una prueba *t* ordinaria, confundiría a aquellos que no fueran estadísticamente sofisticados (como lo era el alumno, antes de leer este capítulo). Asimismo, analizar un experimento con varios grupos utilizando la regresión / correlación múltiples, en lugar del análisis de varianza, confundiría a esos mismos lectores poco sofisticados.

En estos casos, parte de la confusión surge de una cuestión que analizamos en el capítulo 3, y que vale la pena repetir. Muchos confunden la diferencia entre un **diseño de investigación** experimental y correlacional con la diferencia entre métodos estadísticos. Un verdadero diseño experimental de investigación es aquel en el que se asignan individuos al azar a diferentes niveles de la variable de predicción (tales como las condiciones experimental y de control). La asignación aleatoria que mencionamos facilita determinar que los diferentes niveles de la variable de predicción **causaron** algunas diferencias que dieron como resultado la variable dependiente. En el caso de un diseño de investigación correlacional, las variables de predicción e independiente se miden tal como existen. Un ejemplo de ese tipo de diseño sería una encuesta acerca de la relación entre el tiempo compartido y la satisfacción marital. Cuando un estudio del tipo mencionado arroja un

Cuadro 16-2.
Dos mujeres establecen una posición con respecto al sexo y la estadística.

Uno de los textos de estadística avanzada más útiles escritos hasta el momento, es el libro *Utilizando estadística multivariada [Using Multivariate Statistics]*, de Bárbara Tabachnick y Linda Fidell (1996), dos psicólogas de la Universidad del Estado de California en Northridge dedicadas a la investigación. Estas dos mujeres se conocieron durante un almuerzo en la facultad poco después de que Tabachnick fuera contratada. Fidell recuerda que acababa de terminar un curso de francés y otro de álgebra matricial, sólo por el placer de aprender ("En esa época era una persona muy seria"). Se preguntaba qué actividad emprender cuando Tabachnick sugirió que tomaran juntas clases de danzas árabes. Fidell pensó "no está mal algo frívolo para variar". Se equivocaba.

Así comenzó la colaboración entre ellas. Después de las lecciones de danza, mantenían largas discusiones sobre estadística. En particular, descubrieron que compartían la fascinación, y consternación, por los novedosos cálculos estadísticos posibles a través de los nuevos paquetes estadísticos para computadoras. El problema era dar sentido a los resultados.

Fidell describió la situación de la siguiente forma:

> Tenía esta enorme serie de datos para analizar y surgieron una cantidad enorme de bonitos números ordenados en pequeñas y prolijas columnas, pero no estaba segura del significado de todos ellos, e incluso no sabía si mi información había violado algún supuesto crítico. Sabía que existían algunos, pero no sabía nada acerca de ellos. Lo anterior ocurrió en 1975. Yo me había capacitado en la Universidad de Michigan y mis conocimientos sobre estadística abarcaban hasta el análisis de varianza. Pero en esa época, a nadie se le enseñaba el análisis de varianza multivariado. Los paquetes estadísticos que realizaban ese tipo de análisis llegaron posteriormente pero, ¿cómo comprenderlos?

(En el capítulo 17 presentaremos el tema del análisis multivariado de varianza).

Tanto Fidell como Tabachnick habían investigado y aprendido por su cuenta, asistiendo a los cursos necesarios, leyendo, preguntando a otros que conocían mejor los programas, probando qué sucedía si hacían esto o aquello con los datos. Ahora las dos mujeres se preguntaban por qué resultaba todo tan difícil, y si otros estarían volviendo a inventar esa misma rueda en ese mismo momento. Decidieron volcar su propia invención de la rueda en un libro.

"Y así comenzaron quince años de colaboración sin conflictos" informa Fidell. (Hecho que merece compararse con las enemistades narradas en otros cuadros de este mismo libro). Las autoras no tuvieron inconveniente en encontrar un editor, y el libro, que actualmente va por la tercera edición, se ha vendido "muy bien". (A pesar del hecho de que sus títulos preferidos, tales como *El libro de estadística multivariado de Fátima y Scheherazade; Las mil y una variables; El libro rosa borroso de estadística; Pierda peso con la estadística multivariada,* fueron desechados por el editor. Sin embargo, si uno observaba con atención la portada de la primera edición, podía ver una bailarina árabe oculta en el diseño).

Fidell subraya que tanto ella como Tabachnick se consideran analistas de datos y profesoras: no son estadísticas teóricas o prácticas, no han creado métodos, simplemente los han popularizado haciéndolos más accesibles. Sin embargo, pueden nombrar docenas de mujeres que han tenido éxito como estadísticas teóricas. Según Fidell, la estadística es un área en la que la mujer particularmente parece destacarse y sentirse cómoda. Al enseñar a alumnos nuevos, especialmente a aquellos intimidados por la matemática, descubren que por una vez puede hacer que "se relajen", debido a que frecuentemente descubren que disfrutan de la estadística. Ella les dice, "pretendo conquistarlos, y si me dan media oportunidad, lo lograré".

Cualquiera sea la razón, la estadística es una rama de la matemática que, según Fidell, las mujeres con frecuencia consideran "perfectamente lógica, perfectamente razonable, y luego, con el tiempo, algo que realmente pueden disfrutar". Seguramente son buenas noticias para muchas lectoras.

Referencia: Entrevista personal con Linda Fidell.

resultado significativo, sustenta la existencia de una **asociación** entre las dos variables, pero no indica cuál de las variables es causal de la otra (o si alguna tercera variable en común podría estar causando ambas).

Generalmente, los verdaderos diseños experimentales de investigación involucran la asignación a dos o más niveles de la variable de predicción. Tradicionalmente, estos experimentos han sido analizados utilizando una prueba *t* o un análisis de varianza. De hecho, hasta hace poco tiempo atrás, en muchos casos no se enseñaba la regresión / correlación múltiples a los psicólogos experimentales como parte de su capacitación profesional. Eran **experimentalistas** y no debían condescender a la correlación.

Los diseños correlacionales de investigación se utilizan comúnmente cuando no es posible realizar experimentos. Con frecuencia miden la respuesta de determinadas personas con respecto a dos o más variables numéricas, sin tener la posibilidad de que esas personas experimenten completamente una de las variables. (La edad, el nivel de ingresos, el nivel de educación, etc., son ejemplos de variables con las que no es posible poner en práctica la asignación aleatoria). Asimismo, a los sociólogos, economistas y otros científicos sociales no se les enseña la prueba *t* ni el análisis de varianza como parte de su capacitación, debido a que la regresión / correlación es el método apropiado de análisis del cual dependen por completo.

Los diseños experimentales son claramente ventajosos. Por asociación, tanto los diseños como la estadística correlacionales provocan una menor impresión y, fácilmente, se los confunde. Sin embargo, no existe razón para que un verdadero experimento no pueda asignar personas al azar a varios niveles numéricamente diferentes de una variable numérica de predicción. (Utilizamos un ejemplo de este tipo en el capítulo 3, en el que las personas eran asignadas a diferentes cantidades de exposición de una palabra). Un experimento real de ese tipo se analiza adecuadamente sólo con un coeficiente de correlación (y la correspondiente prueba de significación). Si se intentara reducir esos niveles de exposición a dos grupos, por ejemplo la comparación de aquellos con gran cantidad de exposiciones de las palabras con aquellos con poca cantidad de exposiciones, se perdería información y sería un método estadístico más deficiente (entre otros aspectos, el análisis tendría menos potencia).

Asimismo, existen estudios que utilizan diseños correlacionales de investigación en los que, sin embargo, una de las variables tiene sólo dos niveles, como por ejemplo, el género. O podríamos realizar un estudio con *una variable con categorías, con más de dos niveles*, como por ejemplo la nacionalidad. En esos casos, seguramente podríamos analizar los resultados utilizando una prueba t o un análisis de varianza, pero eso no cambiaría el hecho de que los estudios hayan utilizado diseños de investigación correlacionales, en los que resulta difícil discernir la causa y el efecto.

Cabe recalcar que cuando los investigadores seleccionan un método estadístico en lugar de otro, es posible que la decisión esté más relacionada con la costumbre, lo que se "ve bien", e incluso con una confusión, que con cualquier diferencia matemática o lógica entre los procedimientos.

Existe una gran ventaja en utilizar la correlación (o la regresión / correlación múltiples si es necesario) en lugar de la prueba t o el análisis de varianza. El método correlacional proporciona información directa acerca del grado de relación entre la(s) variable(s) de predicción y la variable dependiente, a la vez que permite realizar una prueba de significación. La prueba t y el análisis de varianza sólo brindan la significación estadística. (Si bien con cualquiera de los procedimientos recién mencionados podemos calcular el tamaño del efecto, con un coeficiente de correlación o un coeficiente de correlación múltiple se obtiene automáticamente una indicación del tamaño del efecto con el coeficiente de correlación o regresión en sí mismo).

Otra ventaja de la correlación (y la regresión / correlación múltiples) es que maneja automáticamente el tema de las distintas cantidades de participantes en los grupos que se comparan. Con un análisis de varianza de un criterio, cuando las cantidades de participantes en los grupos son desiguales necesitamos utilizar procedimientos más complicados. Pero, al menos en estos casos, el análisis de varianza de un criterio proporciona resultados precisos.

Por el contrario, al realizar un análisis de varianza de dos o más criterios, si en las casillas hay distintas cantidades de participantes, los procedimientos estándar del análisis de varianza realmente fallan, en el sentido de que su aplicación distorsiona los resultados. En la mayoría de los casos, la mejor solución es replantear el problema a modo de regresión / correlación múltiples.[4]

LOS SUPUESTOS Y EL MODELO LINEAL GENERAL

Otra similitud de las diferentes técnicas basadas en el modelo lineal general es que todos los procedimientos de prueba de hipótesis comparten los mismos supuestos. En el caso de la prueba t y el análisis de varianza, los principales supuestos se refieren a que todas las poblaciones representadas por los grupos tengan la misma varianza y sigan una distribución normal. Los supuestos de las pruebas de significación de correlación, y de regresión / correlación múltiples, son básicamente los mismos, excepto que son un poco más complicado expresarlos.

El supuesto que se refiere a la misma varianza poblacional en la prueba t y en el análisis de varianza coincide en la correlación (y la correlación múltiple) con la igualdad de las varianzas en la parte de la población relacionada con cada nivel de la variable de predicción. Imaginemos un diagrama de dispersión con una recta de regresión. La varianza alrededor de la recta de regresión

[4] La mayoría de los programas para computadoras realizan el proceso mencionado automáticamente cuando se les indica ejecutar un análisis de varianza factorial en el que las cantidades de registros en las casillas no son iguales. Sin embargo, en algunos programas debe darse especialmente la orden para que realicen el proceso mencionado o, de lo contrario, utilizan las fórmulas del análisis de varianza ordinario y arrojan resultados engañosos.

debería ser igual en cada punto a lo largo de la recta. Por ejemplo, supongamos que las observaciones de los niveles inferiores de una variable de predicción tuvieran mucha variación en la variable dependiente, pero que las observaciones en los niveles altos de la variable de predicción tuvieran muy poca variación en la variable dependiente. Esto violaría el principio de igualdad de las varianzas poblacionales. El principio general de igualdad de las varianzas poblacionales, a todos los niveles de una de las variables, se denomina "homoscedasticidad".

En el caso de la correlación y la regresión, el supuesto que se refiere a las distribuciones normales de población se convierte en el requerimiento de que, cada variable, y todas en conjunto, están normalmente distribuidas (lo que se denomina "distribución normal bivariada").

Como hemos visto, todas las técnicas del modelo lineal general arrojan resultados bastante precisos en una amplia gama de situaciones, excepto cuando la cantidad de participantes es muy pequeña o cuando se violan significativamente los supuestos. En verdad, estos distintos métodos constituyen las principales herramientas de la investigación psicológica.

CONTROVERSIAS Y LIMITACIONES

El modelo lineal general, en *sí mismo*, no es muy controvertido; es simplemente una enunciación matemática de una relación entre variables. De hecho, su papel como base de las técnicas estadísticas más importantes aún no ha sido ampliamente comprendido por los investigadores en ejercicio.

Sin embargo, el método de los cuadrados mínimos dentro del modelo lineal general es un poco más controvertido. Una alternativa es minimizar el error absoluto en lugar del error cuadrático. (Una ventaja del método mencionado es que, en lugar de utilizar la raíz cuadrada del promedio de los desvíos cuadráticos como la medida más común de variación, usaríamos simplemente el promedio de los valores absolutos de los desvíos, dando así mucha menos influencia de distorsión a los valores atípicos).

De todos modos, las principales críticas relacionadas con el modelo lineal general son las que involucran la prueba de hipótesis. Son las críticas que hemos estado tratando a lo largo del libro, incluso su carácter robusto por el incumplimiento de los supuestos y la importancia del tamaño del efecto o la prueba de significación.

Existen también críticas en otro sentido, que valen la pena mencionarlas aquí. Se trata de críticas que están relacionadas con el papel que juega la estadística en la ciencia en general, pero que, en la práctica, se formulan más frecuentemente en el contexto de los procedimientos más importantes basados en el modelo lineal general. Se trata de la causalidad. Hemos tratado el tema hasta cierto punto en el capítulo 3 y nuevamente en este capítulo, cuando analizamos el problema de deducir una dirección de causalidad a partir de un estudio que no utiliza asignaciones aleatorias a los distintos grupos. Pero existe una cuestión aún más profunda con respecto a este tema: ¿Cuál es el significado de causalidad?

Baumrind (1983) ha delineado dos interpretaciones de la causalidad que se utilizan en la ciencia. Una, a la que denomina teoría de la causalidad basada en la "regularidad", encuentra sus raíces en filósofos tales como David Hume y John Stuart Mill (al igual que en antiguos científicos, incluyendo a Galileo). Esta perspectiva sostiene que consideramos a X causa de Y si a) X e Y están relacionadas regularmente, b) X precede a Y y c) no existen otras causas anteriores a X que pudieran causar a X y a Y. En psicología, abordamos el punto a buscando una correlación significativa entre X e Y; abordamos el punto b, si es posible, a través de nuestro conocimiento de la situación (en una correlación entre ser el primogénito de una familia y sufrir luego de angustia, podemos excluir la posibilidad de que la angustia sufrida más tarde durante la vida de una persona sea la causa de que esa persona sea primogénita) o diseñando un experimento para averiguarlo (manipulando X antes de medir Y). El punto c) está relacionado con el tema de la correlación entre X e Y, debido a una tercera variable que es causa de las dos primeras. Lo ideal sería abordar el tema a través de la designación aleatoria a los grupos, pero si no es posible solucionar el tema de este modo, se utilizan como estrategia substituta varios métodos estadísticos para igualar a los grupos con respecto a terceros factores propuestos. (En el capítulo 17 analizamos algunas de esas estrategias).

Como psicólogos, sólo a veces nos encontramos en posición de realizar el tipo de investigación experimental rigurosa que nos proporciona una fuerte base para sacar conclusiones con respecto a la causa y el efecto. Así, gran parte de la crítica y de la controversia relacionada con la investigación de aplicación práctica, en la que generalmente es más difícil aplicar métodos rigurosos, frecuentemente gira alrededor de esos temas. Por ejemplo, si el matrimonio y la felicidad están correlacionados, ¿el matrimonio hace más felices a las personas, o las personas felices se casan y permanecen casadas?

Existe otra visión de la causalidad, una visión aún más exigente que considera las condiciones de la teoría de la regularidad como requisitos previos para determinar una causa, pero esas condiciones no son suficientes por sí mismas. Esta segunda visión, a la que Baumrind llama teoría "generativa" de la causalidad, encuentra sus raíces en Aristóteles, Tomás de Aquino e Immanuel Kant. La visión generativa se enfoca en la dinámica con que X afecta Y, el proceso intrínseco por el cual una está conectada con la otra. Es el modo en que interpretan la causalidad la mayoría de las personas no relacionadas con la ciencia (o la filosofía). La idea misma de causalidad puede haber surgido como metáfora de experiencias tales como desear que mi brazo se mueva (**evento** X), y se mueve (**evento** Y). Los científicos también toman muy seriamente esta visión de causalidad, aun si ofrece desafíos mucho más complejos. Se aborda principalmente a través de la teoría y del *análisis cuidadoso de procesos intermedios*. Pero incluso aquellos que recalcan la importancia de esta segunda visión reconocerían que demostrar una conexión confiable entre X e Y (a través de la significación estadística, por ejemplo) es importante, al menos, para identificar los vínculos que requieren investigación para determinar la verdadera conexión causal.

Finalmente, también existen aquellos que sostienen, con algunos buenos argumentos, que demostrar la causalidad no debería ser un objetivo de la psicología científica. Pero ya hemos tenido suficiente controversia para un capítulo.

Resumen

El modelo lineal general equipara el valor de una variable para cualquier individuo con la suma de una constante, más la influencia parcial y ponderada de cada una de otras variables, más el error. El coeficiente de correlación y la regresión / correlación múltiples (y las correspondientes pruebas de significación), la prueba t y el análisis de varianza, son todos casos especiales del modelo lineal general.

La regresión / correlación múltiples es prácticamente idéntica al modelo lineal general, y la regresión y correlación bivariadas son casos especiales de regresión / correlación múltiples, en los que existe sólo una variable de predicción.

La prueba t para medias independientes se puede deducir matemáticamente del análisis de varianza. Es un caso especial del análisis de varianza en el que hay sólo dos grupos. La puntuación t es la raíz cuadrada de la razón F, calculados con los mismos datos. Existen muchas similitudes en las formas de realizar los dos procedimientos: los numeradores de t y

F se construyen sobre las diferencias entre las medias de los grupos; los denominadores de ambos se construyen sobre la varianza interna de los grupos; el denominador de t incluye la división por la cantidad de participantes y el numerador de F incluye la multiplicación por la cantidad de participantes; y los grados de libertad de t son iguales a los grados de libertad del denominador de F.

La prueba t para medias independientes también es un caso especial de la prueba de significación del coeficiente de correlación. Una correlación mide el grado de relación de una variable de predicción o independiente con una variable dependiente. Del mismo modo, al indicar la diferencia entre las medias de los grupos, la prueba t identifica una relación entre la variable cuyos grupos están divididos, es decir, la variable independiente o de predicción con la variable dependiente. Si asignamos el valor 1 a cada participante en uno de los dos grupos y el 2 a cada participante en el otro grupo (o dos números diferentes cualesquiera), y después calculamos una correlación de esos valores con la variable dependiente, la significación de la correlación será igual que la producida por la prueba t. Si dibujamos un diagrama de dispersión con los datos mencionados obtendríamos una columna de valores observados para cada grupo, y la línea de regresión pasaría por las medias de cada uno de ellos. Cuanto más diferentes sean las medias, mayor será la reducción proporcional del error con respecto a utilizar la gran media, y mayor será la puntuación basada en una comparación de las medias de los dos grupos.

El análisis de varianza y la correlación / regresión también presentan muchas similitudes. SC_{Total}, en la regresión y en el análisis de varianza, se refiere a los desvíos de cada observación con respecto a la media de todas las observaciones de la variable dependiente. Las medias grupales en un análisis de varianza son los valores predichos para cada individuo en la regresión; así, SC_{Error} y SC_{Dentro} son iguales. La reducción de error cuadrático ($SC_{Total} - SC_{Error}$) en la regresión es igual a la suma de los desvíos cuadráticos de las medias de los grupos de observaciones con respecto a la gran media ($SC_{Intergrupal}$) en el análisis de varianza. Finalmente, la reducción proporcional del error de la regresión (r^2 o R^2), en la regresión, es igual a la proporción de varianza explicada (R^2 o eta^2) por el tamaño del efecto en el análisis de varianza.

Todo análisis de varianza puede plantearse como una regresión múltiple, transformando las categorías que representan los diferentes grupos en una o más variables numéricas dicotómicas. En sentido estricto, el análisis de varianza es un caso especial de regresión múltiple, en el que las variables de predicción se establecen del modo descripto precedentemente.

Todos los métodos mencionados comparten los mismos supuestos en cuanto a que las distribuciones de la población son normales y tienen igual varianza en todos los niveles de la variable de predicción.

La prueba t, el análisis de varianza y la correlación pueden plantearse como regresión / correlación múltiples; sin embargo, la práctica convencional hace que estos procedimientos conceptualmente idénticos se utilicen en diferentes contextos de investigación, como si en realidad fueran diferentes.

Con respecto a la causalidad, la teoría de la regularidad identifica a X como causa de Y, si X e Y están relacionadas, X precede a Y, y no existe un tercer factor que preceda a X y pudiera causar ambas. La teoría generativa sostiene que, además, debe comprenderse claramente el mecanismo por el cual X afecta Y. Los procedimientos estadísticos pueden demostrar una relación entre X e Y, e incluso a veces pueden contribuir con evidencia contra una tercera variable propuesta como causa de X e Y. Toda otra prueba de que X sea causa de Y depende del conocimiento de la situación, del diseño experimental y del análisis teórico.

Términos clave

- Modelo lineal general.
- Modelo de cuadrados mínimos.
- Codificación nominal.

Ejercicios

Los ejercicios implican la realización de cálculos (con la ayuda de una calculadora). La mayoría de los problemas estadísticos reales se resuelven por computadora, pero aunque exista la posibilidad de utilizarla, es conveniente realizar estos ejercicios manualmente para incorporar el método de trabajo.

Para adquirir práctica en la utilización de una computadora, para resolver problemas estadísticos, se puede utilizar la sección de computación de cada capítulo, publicada en la *Guía de estudio y libro de tareas de computación para el alumno [Student's Study Guide and Computer Workbook]* que acompaña este libro.

Todos los datos de esta sección son ficticios (a menos que se especifique lo contrario).

Las respuestas a los ejercicios de la serie I se encuentran al final del libro.

SERIE I

1. a) Busque y anote el punto de corte t al nivel 0,05 (dos colas), correspondiente a 5, 10, 15 y 20 grados de libertad. b) Eleve cada punto de corte t al cuadrado y anótelo al lado del t. c) Busque y anote, al lado de los t cuadráticos, los puntos de cortes para distribuciones F con 1 grado de libertad en el numerador y 5, 10, 15 y 20 grados de libertad como denominadores. (Los resultados deberían ser iguales, con diferencias de redondeo).

2. A continuación aparecen tres series de datos. En el caso de las primeras dos series de datos, además de las medias y las varianzas poblacionales estimadas, incluimos la información correspondiente a la prueba t. En el tercer caso debe calcular usted mismo esa última información. Además, para cada caso, calcule un análisis de varianza de un criterio. Observe las similitudes entre a) el gl de t y el gl del denominador de F, b) el punto de corte t y la raíz cuadrada del punto del corte F, c) $S^2_{\text{Combinada}}$ y S^2_{Dentro} y d) el valor t y la raíz cuadrada de la razón F. t y F corresponden al nivel 0,05; las pruebas t son de dos colas.

	Grupo experimental			Grupo control			Prueba t			
	N	M	S^2	N	M	S^2	gl	t necesario	$S^2_{combinada}$	t
(i)	30	12,0	2,4	30	11,1	2,8	58	2,004	2,6	2,16
(ii)	36	100	40	36	104	48	70	1,995	44	2,56
(iii)	16	73	8	16	75	6				

3. El grupo A está formado por 10 personas, cuyos valores observados presentan una media de 170 y una estimación de la varianza de 48. El grupo B también está formado por 10 personas: $M = 150$, $S^2 = 32$. Realice una prueba t para medias independientes (dos colas) y un análisis de varianza; calcule los dos procedimientos en las mitades de una misma página y disponga los cálculos de forma paralela uno al lado del otro. (Es decir, cree una tabla similar, en cuanto a diseño, en la parte inferior de la tabla 16-2). Utilice el nivel 0,05.

4. Calcule un análisis de varianza con los valores que se enumeran a continuación; luego calcule un análisis de regresión, incluyendo el correspondiente diagrama de dispersión (y la recta de regresión) indicando el coeficiente de correlación (entre el grupo en el que se encuentran los sujetos y sus valores observados), determinando la reducción proporcional del error a través del método extenso en el que se calculan los valores predichos y se determina el error cuadrático medio utilizando esos valores; por último, confeccione un cuadro que indique las coincidencias de los resultados.

Grupo A	Grupo B
13	11
16	7
19	9
18	
19	

5. Con los valores que se enumeran a continuación, calcule una prueba t para medias independientes (utilizando el nivel 0,05, una cola), el coeficiente de correlación (entre el grupo en el que se encuentran los participantes y sus valores observados) y el t para la significación del coeficiente de correlación (utilizando la fórmula $t = r\sqrt{N-2} / \sqrt{1-r^2}$).

Grupo A	Grupo B
0,7	0,6
0,9	0,4
0,8	0,2

6. Explique los vínculos principales entre la regresión múltiple y el análisis de varianza.

SERIE II

1. a) Busque y anote el punto de corte F al nivel 0,01 para distribuciones con 1 grado de libertad en el numerador y 10, 20, 30 y 60 grados de libertad en el denominador. b) Calcule la raíz cuadrada de cada uno y anótela al lado del corte. c) Busque los puntos de corte de la distribución t al nivel 0,01 (dos colas), utilizando, 10, 20, 30 y 60 grados de libertad, y anótelos al lado de las correspondientes raíces cuadradas de F. (Los resultados deberían ser idénticos, teniendo en cuenta las diferencias de redondeo).

2. A continuación enumeramos tres series de datos, todos ellos tomados del ejercicio 2, serie II, del capítulo 10. Si no calculó antes las pruebas t para estos datos, hágalo ahora, pero esta vez utilizando el nivel 0,01 de dos colas. Luego, en cada caso, calcule además un análisis de varianza de un criterio (también al nivel 0,01). Observe las similitudes entre a) el gl de t y el gl del denominador de F, b) el punto de corte t y la raíz cuadrada del punto de corte F, c) $S^2_{Combinada}$ y CM_{Dentro} y d) el valor t y la raíz cuadrada de la razón F.

	Grupo experimental			Grupo control		
	N	M	S^2	N	M	S^2
(i)	10	604	60	10	607	50
(ii)	40	604	60	40	607	50
(iii)	10	604	20	10	607	16

3. Realice una prueba t para medias independientes (dos colas) y un análisis de varianza con los valores que aparecen a continuación. Realice los cálculos en las mitades de una misma página y coloque los cálculos paralelamente, uno al lado del otro. (Es decir, cree una tabla similar, en cuanto al diseño, en la parte inferior de la tabla 16-2). Utilice el nivel 0,01.

Grupo A	Grupo B
0	4
1	5
0	6
	5

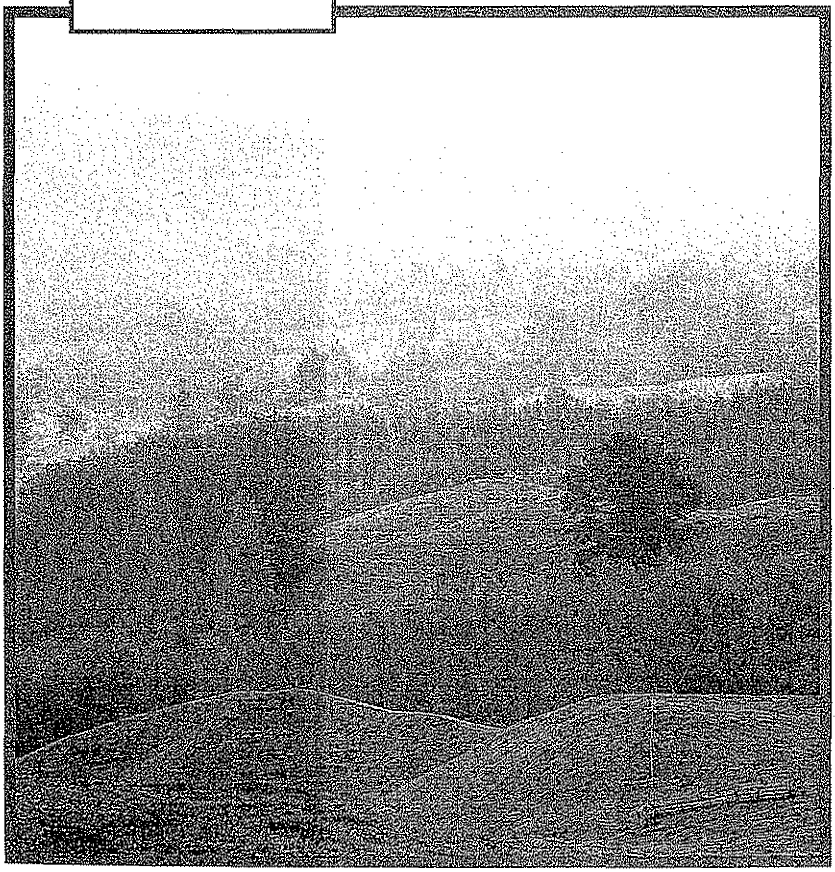
4. Con los datos del ejercicio 3 (anterior), calcule un análisis de regresión, incluya un diagrama de dispersión y calcule el coeficiente de correlación (entre el grupo en el que se encuentran los participantes y sus valores observados), determinando la reducción proporcional del error a través del método extenso en el que se calculan los valores predichos y luego se determina el error cuadrático medio utilizando esos valores; después calcule la significación de la correlación (utilizando la fórmula $t = r\sqrt{N-2} / \sqrt{1-r^2}$ y elevando luego t al cuadrado). Finalmente, realice un cuadro que muestre las coincidencias de los resultados.

5. Con los valores **enumerados a continuación**, calcule a) una prueba t para medias independientes (nivel 0,05, dos colas); b) el coeficiente de correlación (entre el grupo en el que se encuentran los participantes y sus valores observados) y el t para la significación del coeficiente de correlación ; c) un análisis de varianza (nivel 0,05), y d) el coeficiente chi-cuadrado y phi (para el chi-cuadrado, cree una tabla de contingencia 2 x 2 en la cual una dimensión sean los grupos y la otra dimensión se refiera al valor 0 ó 1 de los participantes en la medida dependiente). e) Confeccione una tabla con las similitudes de los cálculos y los resultados.

6. Explique la codificación nominal.

Group A	Group B
0	0
0	0
0	0
0	0
0	0
0	0
0	1
0	1
0	1
0	1
0	1
0	1
1	1
1	1
1	1
1	1

17 Comprensión de los procedimientos estadísticos avanzados que aparecen en publicaciones científicas



Descripción del capítulo

- Breve revisión de correlación y regresión múltiples.
- Regresión múltiple jerárquica y por pasos.
- Correlación parcial.
- Confiabilidad.
- Análisis factorial.
- Modelo causal.
- ANCOVA (*Analysis of Covariance*, Análisis de covarianza).
- MANOVA (*Multivariate Analysis of Variance*, Análisis de varianza multivariado) y MANCOVA (*Multivariate Analysis of Covariance*, Análisis de covarianza multivariado).
- Repaso general de técnicas estadísticas.
- Controversia: ¿Debería ser controvertida la estadística?
- ¿Cómo leer resultados en publicaciones científicas que incluyen técnicas estadísticas que no nos son familiares?
- Resumen.
- Términos clave.
- Ejercicios.

La mayoría de las investigaciones leídas por alumnos de psicología utilizan uno o más de los procedimientos que hemos aprendido a través de este libro. Sin embargo, a veces aparecerán procedimientos que no se enseñan sino en cursos de estadística más avanzados. Afortunadamente, esos procedimientos son, por lo general, extensiones directas de lo que ya hemos aprendido; tal vez no tan directas como para entender todos sus detalles y limitaciones, pero sí para poder comprender la idea general de lo que se está realizando con los datos resultantes del estudio.

Podemos dividir esos procedimientos estadísticos avanzados en aquellos que se concentran en las asociaciones entre variables y aquellos que se concentran en las diferencias entre los grupos (aunque, tal como hemos aprendido en el capítulo 16, esa distinción es algo artificial). Los procedimientos que trataremos primero se concentran en asociaciones entre variables. Todos ellos son básicamente extensiones y elaboraciones de lo aprendido en los capítulos 3 y 4 sobre correlación y regresión. Después de una breve revisión de la regresión múltiple como base de los demás procedimientos, presentamos las regresión múltiple jerárquica y la gradual, la correlación parcial, la confiabilidad, el análisis factorial y el modelo causal. Luego nos abocamos a las técnicas que se concentran en las diferencias entre grupos. Se trata básicamente de extensiones o elaboraciones de lo aprendido en los capítulos 11 a 13 sobre el análisis de varianza. Incluimos en este grupo a los procedimientos de análisis de covarianza, análisis de varianza multivariado y análisis de covarianza multivariado. La controversia del capítulo cuestiona si la estadística debería ser controvertida. Finalmente, concluimos el capítulo con una exposición acerca de cómo actuar cuando encontramos publicaciones científicas que utilizan técnicas estadísticas que no conocemos.

BREVE REVISIÓN DE LA CORRELACIÓN Y LA REGRESIÓN MÚLTIPLES

En el capítulo 4 hemos aprendido la correlación y la regresión múltiples (y en el capítulo 16 hemos repasado esas técnicas brevemente). La correlación múltiple se basa en la relación de una variable dependiente con la combinación de dos o más variables de predicción. En un ejemplo ficticio que utilizamos en el capítulo 4, existía una correlación múltiple (R) de 0,96 entre el nivel de estrés experimentado por varios gerentes y la combinación de la cantidad de empleados que supervisaban, el nivel de ruido en el lugar de trabajo y la cantidad de decisiones que debían tomar cada mes.

También aprendimos que la regresión múltiple es la predicción de una variable dependiente sobre la base de dos o más variables de predicción. (Cabe recordar que la regresión es simplemente la forma de predicción de la correlación). La reglas de predicción de la regresión múltiple presenta un coeficiente de regresión para cada variable de predicción. Si se conoce el valor observado de una persona en las variables de predicción, se multiplica ese valor de cada variable de predicción por el coeficiente de regresión de esa variable. La suma de los productos será el valor predicho para esa persona en la variable dependiente. Cuando se trabaja con puntuaciones Z, los coeficientes de regresión son coeficientes de regresión estandarizados, denominados valores ponderados beta (β). Por ejemplo, con tres variables independientes, la ecuación de la regla de predicción es la siguiente:

$$\hat{Z}_Y = (\beta_1)(Z_{X_1}) + (\beta_2)(Z_{X_2}) + (\beta_3)(Z_{X_3})$$

En el ejemplo del estrés sufrido por los gerentes, la regla de predicción de regresión múltiple con puntuaciones Z era la siguiente:

$$\hat{Z}_{\text{Estrés}} = (0{,}51)(Z_{\text{Empleados}}) + (0{,}11)(Z_{\text{Ruido}}) + (0{,}33)(Z_{\text{Decisiones}})$$

Cuando se trabaja con puntuaciones originales, el coeficiente de regresión para puntuaciones originales (b) se multiplica por la puntuación original en cada variable de predicción, y se suma la constante de regresión de puntuaciones originales (a). La siguiente es la fórmula con tres variables independientes:

$$\hat{Y} = a + (b_1)(X_1) + (b_2)(X_2) + (b_3)(X_3)$$

En el ejemplo del estrés de los gerentes, la regla de predicción de regresión múltiple con puntuaciones originales era la siguiente:

$$\hat{Y} = -4{,}70 + (0{,}56)(\text{Empleados}) + (0{,}6)(\text{Ruido}) + (0{,}86)(\text{Decisiones})$$

En la correlación y regresión múltiples, los investigadores pueden determinar la significación estadística tanto del coeficiente de correlación múltiple general (R) como de cada beta individual (ó b). En la mayoría de los casos, si el R general no es significativo, el investigador no prueba la significación de los coeficientes betas individuales. Sin embargo, es bastante probable que el coeficiente R general sea significativo pero que algunos de los coeficientes betas no lo sean. Por ejemplo, la correlación significativa general podría ser el resultado de la fuerte influencia de una sola variable de predicción, con leves contribuciones por parte de las otras variables.

Regresión múltiple jerárquica

A veces, los investigadores están interesados en observar la influencia de diversas variables de predicción en forma secuencial. Es decir, les interesa saber, en primer lugar, cuál es la correlación entre la primera variable de predicción y la variable dependiente. En segundo lugar, les interesa saber cuánto agrega a la correlación múltiple general el hecho de incluir una segunda variable de predicción. Luego, el investigador puede estar interesado en saber cuánto más se agrega incluyendo una tercera variable de predicción, y así sucesivamente. El procedimiento que acabamos de describir es conocido con el nombre de **regresión múltiple jerárquica**.

En las investigaciones que utilizan la regresión múltiple jerárquica, la cantidad agregada sucesivamente por cada variable a la predicción general se describe usualmente en términos de un aumento de R^2 (proporción de varianza explicada).

Analicemos un ejemplo. MacDonald y sus colegas (1997) analizaron la relación de varios factores con los PTSD (*Posttraumatic stress disorder*, Trastornos de estrés postraumático) de veteranos de guerra. Los psicólogos reclutaron una muestra comunitaria de 756 veteranos de la Guerra de Vietnam oriundos de Nueva Zelanda, que incluía 161 maoríes (los maoríes son el pueblo polinesio autóctono de Nueva Zelanda). La tabla 17-1 indica los resultados del análisis de regresión jerárquica realizado por los psicólogos. La primera variable que se tomó en cuenta (paso 1) era la exposición al combate, que por sí misma tenía una R^2 de 0,07; los dos asteriscos indican que el resultado es significativo. El segundo paso fue agregar una serie de variables relacionadas con la experiencia militar en Vietnam. La serie completa de variables aumentó el R^2 de 0,07 a 0,171, un aumento (cambio en R^2) de 0,101. (En la tabla, los investigadores señalan un aumento de 0,100. La diferencia probablemente se deba al redondeo de las cantidades indicadas en la tabla para cada etapa).

Tabla 17-1.
Coeficientes de regresión, valores R^2 y cambio en R^2, correspondientes a la exposición al combate, las variables de experiencias en Vietnam y la raza, que predicen el PTSD.

	Coeficiente beta estandarizado		
Variable de predicción	*Paso 1*	*Paso 2*	*Paso 3*
Exposición al combate	0,266***	0,300**	0,297**
Exposición militar en Vietnam			
Tiempo de servicio en Vietnam		−0,035	−0,036
Rango		−0,316**	−0,314**
Rol en combate		0,153*	0,154*
Especialización militar 1[a]		0,015	0,017
Especialización militar 2[a]		0,044	0,044
Raza[b]			−0,024
R^2	0,070***	0,171**	0,171**
Cambio en R^2		0,100*	0,001

$*p < 0{,}01$; $**p < 0{,}001$.
[a]Variable nominal: especialización militar 1 (infantería/no infantería); especialización militar 2 (artillería / no artillería).
[b]Variable dicotómica (maorí/no maorí).
Fuente: MacDonald, C., Chamberlain, K., & Long, N. (1997), tab. 2. "Raza, combate y estrés postraumático [PTSD] en una muestra comunitaria de veteranos de la Guerra de Vietnam oriundos de Nueva Zelanda". *Revista científica de estrés traumático* [*Journal of Traumatic Stress*], 10, 123. Copyright, 1997, por la Sociedad Internacional de Estudios sobre Estrés Traumático. Reimpreso con autorización.

Finalmente, en el paso 3 se agregó la raza (maorí o no). La inclusión de la variable raza no aumentó mucho el R^2.

Los resultados son especialmente interesantes debido a que la correlación bivariada directa entre la raza y los PTSD era significativa. Aun así, la regresión jerárquica muestra que la raza no predice el estrés postraumático si ya se ha tenido en cuenta la experiencia en combate y la situación militar. En otras palabras, el efecto aparente de la raza probablemente se debía a que la exposición al combate y la experiencia militar de los maoríes eran diferentes.

Analicemos otro ejemplo. Hermann y sus colegas (1997) realizaron un estudio de niños bajo tratamiento de biorretroalimentación de la migraña. El objetivo del estudio era identificar variables que predijeran el éxito de ese tipo de tratamiento y, para eso, los investigadores midieron la AJ (Actividad de las jaquecas) al finalizar un programa de 8 semanas de tratamiento. La AJ era una variable que resumía factores tales como la frecuencia y la intensidad de los dolores de cabeza. Los investigadores también midieron una cantidad de variables de predicción, que incluían la AJ "preexistente" (antes de que comenzara el tratamiento), los problemas de conducta según la CBCL (*Child Behavior Checklist*, Lista de verificación del comportamiento infantil), incluyendo la interiorización y exteriorización del comportamiento, las aflicciones psicosomáticas, la edad y las variables familiares según la FES (*Family Environment Scale*, Escala de ambiente familiar), incluso el índice de relaciones familiares, la organización y el control.

Los investigadores utilizaron la regresión jerárquica para observar si las demás variables de predicción agregaban algo a la predicción de la efectividad por encima de la AJ preexistente. En realidad, realizaron dos regresiones jerárquicas diferentes. Una, a la que llamaron modelo 1, examinaba si las variables referidas al niño (comportamiento y edad) aportaban alguna contribución por encima de la AJ preexistente. La otra, el modelo 2, examinaba si las variables familiares aportaban alguna contribución por encima de la AJ previa a la línea de base. Los investigadores informaron sus resultados de la siguiente forma:

Tabla 17–2
Predicción del efecto de determinado tratamiento a través de la regresión múltiple jerárquica.

Modelo y paso	R^2 Total	*gl*	*F*	β	$R^2\Delta$	$F\Delta$	
Modelo 1	0,39	5, 26	3,3*				
Paso 1. AJ previa a la línea de base				−0,04	0,02	0,7	
Paso 2. CBCL: interiorización del comportamiento					−0,27	0,37	3,9*
CBCL: exteriorización del comportamiento					0,57*		
Aflicciones psicosomáticas				0,37*			
Edad				−0,43*			
Modelo 2	0,03	4, 27	0,2				
Paso 1. AJ previa a la línea de base				−0,18	0,02	0,7	
Paso 2. FES: índice de relaciones familiares				−0,04	0,01	0,06	
FES: organización				−0,06			
FES: control				−0,05			

Nota: AJ = actividad de las jaquecas; CBCL = Lista de verificación del comportamiento infantil; FES = Escala de ambiente familiar.

*$p < 0,05$.

Fuente: Hermann, C., Blanchard, E. B., & Flor, H. (1997), tab. 5. "Tratamiento de biorretroalimentación para la migraña: predicción del efecto del tratamiento". *Revista científica de psicología de asesoramiento y clínica [Journal of Consulting and Clinical Psychology]* 65, 611–616. Copyright, 1997, por la Asociación Americana de Psicología. Reimpreso con autorización.

Utilizando la regresión múltiple jerárquica, se evaluaron independientemente las características del niño (modelo 1) y el ambiente familiar (modelo 2) como factores de predicción del efecto del tratamiento. Para controlar las diferencias de la línea de base, se ingresó primero la AJ previa a la línea de base. Las variables que reflejan las características del niño y los aspectos del funcionamiento familiar respectivamente, fueron ingresados como series en el paso 2 . . . (pp. 613-614)

Regresión múltiple por pasos

Con frecuencia, especialmente en un estudio exploratorio, el investigador puede haber medido muchas variables de predicción potenciales y, además, necesitar seleccionar aquellas que aportan una contribución útil a la predicción general. Por lo general, esto se logra con la **regresión múltiple por pasos.** La forma más común de la regresión múltiple gradual funciona del siguiente modo: un programa de computación realiza un proceso paso a paso, seleccionando primero la variable que presenta la mayor correlación con la variable dependiente. Si esa correlación no es significativa, el proceso se detiene, ya que, incluso el mejor elemento de predicción no es útil. Pero si esa primera correlación es significativa, el proceso continúa. El siguiente paso es seleccionar la variable de predicción que, en combinación con la primera, presenta el *R* más elevado. Entonces, la computadora verifica si la combinación implica una mejora significativa por encima de la mejor variable de predicción independiente. Si no lo es, el proceso se detiene. Si realmente implica una mejora significativa, el programa continúa. El siguiente paso es seleccionar cuál de las restantes variables de predicción al analizarse en combinación con las dos primeras, crea el *R* múltiple más elevado. Luego se verifica la combinación para observar si implica una mejora significativa para la predicción por encima de los dos primeros factores de predicción. El proceso continúa hasta que se incluyen todas las variables de predicción o hasta que el punto en el cual agregar cualquiera de las restantes no implica ninguna mejora significativa. La tabla 17-3 muestra un diagrama del procedimiento descripto, al que se denomina "por pasos", porque avanza de a un paso a la vez.[1] El siguiente es un ejemplo tomado del estudio que acabamos de analizar acerca de los elementos de predicción de la efectividad de un tratamiento de biorretroalimentación para niños con migraña. Además de las regresiones jerárquicas, Hemann et al. también realizaron una regresión gradual. El informe es el siguiente:

> Se calculó una regresión gradual [...] con fines exploratorios, comparando en forma directa todas las variables de predicción entre sí, excepto la AJ previa a la línea de base. Coincidentemente con el modelo 1, la edad ($\beta = 0{,}38$) y las aflicciones psicosomáticas ($\beta = 0{,}39$) resultaron variables exitosas para la predicción del efecto del tratamiento, justificando el 35% del mismo, $F(3, 28) = 4{,}9$, $p<0{,}01$ (p. 614).

Lo que los autores quieren decir es que, de las diversas variables de predicción (que eran siete), la proporción de varianza justificada por dos de ellas (edad y aflicciones psicosomáticas) no era aumentada significativamente al incluir también cualquiera de las variables restantes.

[1] Técnicamente, lo descripto es una "regresión por pasos hacia adelante". Algunos investigadores prefieren comenzar con una regla de predicción que incluya todas las variables de predicción para observar luego cuánta capacidad de predicción se pierde al eliminar el factor de predicción menos útil. Si no se pierde mucha capacidad de predicción, se elimina la siguiente variable menos útil, y así sucesivamente. El proceso continúa hasta que queda una pequeña serie de variables cuyo poder de predicción se reduce significativamente al eliminar el factor de predicción menos útil. Este procedimiento alternativo se denomina "regresión por pasos hacia atrás". En la mayoría de los casos, las regresiones por pasos hacia adelante o hacia atrás producen aproximadamente los mismos resultados; realmente, la utilización de uno y otro proceso no denota gran diferencia sino que se trata más bien de una cuestión de preferencia del investigador.

Tabla 17-3.
Proceso de regresión múltiple gradual.

Paso 1:	Buscar todas las variables de predicción potenciales y encontrar la mejor correlación bivariada con la variable dependiente.	
Paso 2:	Probar la significación.	
	Si la correlación no es significativa, →	DETENERSE
	Si la correlación es significativa, incluir esa variable en todos los pasos siguientes y →	CONTINUAR
Paso 3:	Buscar entre todas las restantes variables potenciales de predicción aquélla que combine mejor con las ya incluidas para predecir la variable dependiente.	
	Si ninguna otra resulta significativa, →	DETENERSE
	Si existe una que es significativa, incluirla en todos los siguientes pasos y →	REPETIR EL PASO 3 PARA ENCONTRAR LA MEJOR VARIABLE DE PREDICCIÓN ENTRE LAS RESTANTES

Es preciso hacer una advertencia en cuanto a la regresión gradual: la fórmula de predicción resultante es el grupo de variables que mejor predice la variable dependiente, **basándose en la muestra analizada.** Sin embargo, sucede con frecuencia que cuando se analizan las mismas variables con una nueva muestra, la mejor combinación de variables resulta ser, en cierto modo, diferente.

Comparación entre la regresión jerárquica y regresión por pasos

Las regresiones jerárquicas y graduales son similares en aspectos importantes. Con ambos métodos se agrega una o más variables a la vez y se verifica si lo que se ha agregado aumenta significativamente la predicción. Sin embargo, también existe una diferencia importante: en la regresión jerárquica, el orden en que se agregan las variables de predicción se basa en algún plan o teoría, establecido de antemano por el investigador. En cambio, en la regresión por pasos no existe ningún plan inicial; la computadora simplemente calcula cuáles son las variables que más conviene agregar, hasta llegar al punto en el que agregar más variables no implica ninguna contribución adicional.

Así, la regresión jerárquica se utiliza en investigaciones basadas en una teoría o en algún conocimiento sustancial previo, mientras que la regresión por pasos es útil en los estudios exploratorios en los que no existe un resultado esperado, o bien, en investigaciones aplicadas en las que se busca la mejor fórmula de predicción sin importar su significado teórico.

CORRELACIÓN PARCIAL

La **correlación parcial** es otra técnica ampliamente utilizada en la psicología de la personalidad y del desarrollo, en la psicología clínica y social y en varias otras áreas aplicadas a la psicología. La correlación parcial es el grado de asociación entre dos variables, por encima de la influencia de otra u otras variables. Supongamos que un investigador necesita saber hasta qué punto el estrés sufrido por una persona en la vida marital está relacionado con el tiempo que esa persona ha estado casada. Sin embargo, el investigador es consciente de que parte de lo que podría relacionar al estrés marital con el tiempo de casado, es que las personas que llevan más tiempo casadas probablemente tengan hijos, y ese hecho podría causar estrés marital. Por lo tanto, calcular simplemente la correlación entre el estrés marital y el tiempo de matrimonio sería engañoso; lo que el investigador necesita saber es la relación que existiría entre el estrés y el tiempo de matrimonio si todas las parejas tuvieran la misma cantidad de hijos. O, para decirlo de otra manera, el investigador necesita que, de algún modo, la información derivada del estrés y el tiempo de matrimonio no incluya lo aportado por la cantidad de hijos de ese matrimonio. Lo anterior se logra mediante la correlación parcial.

En el caso mencionado anteriormente, el investigador calcularía una correlación parcial entre el estrés marital y el tiempo de matrimonio **manteniendo constante** la cantidad de hijos. El procedimiento también se describe como **exclusión** o **control** de la cantidad de hijos (los términos "mantener constante", "excluir" y "controlar" tienen el mismo significado y pueden utilizarse indistintamente). El cálculo estadístico real de la correlación parcial se denomina **coeficiente de correlación parcial**. Este presenta valores desde el -1 al $+1$ y se considera igual a una correlación común entre dos variables, excepto por el hecho de que existe una tercera variable que está siendo controlada.

El siguiente es otro modo de ver la correlación parcial: en el ejemplo que hemos estado mencionando, el investigador podría calcular la correlación entre el estrés y el tiempo de matrimonio utilizando sólo personas que no tuvieran hijos; luego podría calcular la misma correlación sólo con aquellos que tienen un sólo hijo, y así sucesivamente. Cada una de esas correlaciones analizadas independientemente no se ven afectadas por las diferentes cantidades de hijos, ya que entre las personas estudiadas en cada una de las correlaciones no existe esa diferencia. Después, el investigador podría calcular algún tipo de promedio entre las diferentes correlaciones, ninguna de las cuales ha sido afectada por la cantidad de hijos. El promedio entre esas correlaciones es la correlación parcial. Se trata literalmente de una correlación que mantiene una cantidad constante de hijos.

En realidad, los cálculos de una correlación parcial son bastante directos, y no es necesario realizar todas esas correlaciones individuales ni el promedio de ellas. Sin embargo, el resultado del proceso es el mismo que si se realizaran esos cálculos. [2]

La correlación parcial, en líneas generales, se utiliza para seleccionar una de varias explicaciones teóricas alternativas de las relaciones entre variables. Supongamos que un investigador descubre una correlación común entre el estrés marital y el tiempo de matrimonio, y está interesado en utilizar ese resultado para sustentar la teoría de que el paso del tiempo hace que las personas se sientan más estresadas con respecto al matrimonio, ya que cada miembro de la pareja da por sentado al otro. Sin embargo, el investigador también es consciente de que otra explicación posible sería que cuando las personas llevan más tiempo de casadas, probablemente tienen más hijos, y el hecho de tener hijos podría crear estrés en el matrimonio. Si se descubre una correlación entre el estrés y el tiempo de matrimonio, aun después de controlar la cantidad de hijos, la última explicación alternativa referida a la cantidad de hijos se torna improbable.

[2] La correlación parcial está muy relacionada con la regresión múltiple. Por ejemplo, un coeficiente de regresión indica en qué medida una variable en particular es adecuada para predecir la variable dependiente, dado cualquier nivel de todas las otras variables de la ecuación. Además, en la regresión múltiple jerárquica, la contribución que surge al agregar una variable a aquellas ya incluidas en la ecuación, indica, en efecto, lo que esa variable aporta independientemente de todas las demás. (El nombre formal de lo que una variable aporta en una regresión múltiple jerárquica es la "correlación semiparcial"). Para comprender de manera general una publicación científica, digamos que la correlación parcial, el coeficiente de regresión y la cantidad aportada por una variable en la regresión jerárquica, indican algo similar: la relación entre dos variables independientemente de otra u otras variables.

Tabla 17-4.
Correlación bivariada y parcial que indican las relaciones entre las escalas de represión e inseguridad y las preocupaciones autocríticas e interpersonales.

	r **Bivariada**		*r* **Parcial**	
Escala	*Interpersonal*	*Autocrítica*	*Interpersonal*	*Autocrítica*
Represión				
Padres	0,12	0,23***	0,00	0,18**
Madres	0,08	0,23***	–0,12*	0,14*
Inseguridad				
Padres	0,24***	0,13	0,20**	0,02
Madres	0,33***	0,12*	0,29***	–0,07

Nota: Análisis de correlaciones parciales que evalúan las relaciones entre la represión (o inseguridad) y las inquietudes depresivas con control de la inseguridad (o represión) y de la depresión adolescente.

*$p < 0,05$; **$p < 0,01$; ***$p < 0,001$.

Fuente: Frank, S. J., Poorman, M. O., & Van Egeren, L. A. (1997), tab. 5. "Percepción con respecto a las relaciones con sus propios padres por parte de adolescentes internados con preocupaciones depresivas y estado de depresión". *Revista científica de psicología clínica infantil [Journal of Clinical Child Psychology]*, 26, 205–215. Copyright © 1997 por Lawrence Erlbaum Associates, Inc. Reimpreso con autorización.

El siguiente ejemplo fue tomado de una investigación real: Frank y sus colegas (1997) realizaron un estudio de las inquietudes depresivas de adolescentes y su relaciones con sus propios padres.

Los investigadores se concentraron en dos aspectos de las inquietudes depresivas: la preocupación autocrítica y la preocupación interpersonal. También se concentraron en dos aspectos de lo que denominaron "conflicto de separación-individuación" con los padres, es decir, en qué medida los adolescentes percibían represión por parte de sus padres (ejerciendo un fuerte control sobre sus comportamientos) y hasta qué punto los adolescentes se sentían inseguros con respecto a sus padres. Frank et al. informaron el resultado de los análisis de la siguiente forma:

> Después correlacionamos las escalas de percepción -inseguridad y represión, referidas a los padres y las madres, con los valores de inquietudes autocríticas e interpersonales. La tabla [17-4] resume los análisis de correlación bivariada y parcial. En los análisis parciales se controló un aspecto del conflicto de separación-individuación [....] y cada tipo de inquietud depresiva.
> Aunque la magnitud de las correlaciones no es tan amplia, el patrón general de los resultados que muestra [la] tabla es significativo. Los adolescentes que percibían que los padres reprimían sus intentos de separación presentaban más probabilidades de tener inquietudes autocríticas, mientras que los adolescentes que reconocían que ellos mismos estaban experimentando temores y ansiedades con respecto a la separación presentaban más probabilidades de preocupares por inquietudes interpersonales (p. 211).

CONFIABILIDAD

Es poco común que, en psicología, las medidas sean perfectamente precisas. (Tratamos brevemente el tema en el capítulo 3 y lo analizamos con más detalle en el apéndice A). El grado de coherencia o estabilidad de una medición se denomina **confiabilidad.** En líneas generales, la

confiabilidad implica hasta qué punto se obtendría el mismo resultado si se hiciera la misma medición nuevamente a la misma persona bajo las mismas circunstancias. Calcular la **confiabilidad** de un procedimiento de medición es un tema clave en casi todas las áreas de investigación psicológica, sin importar si los procedimientos son cuestionarios, entrevistas, observaciones de comportamientos, reacciones fisiológicas u otros. Los cálculos estadísticos de la confiabilidad aparecen con frecuencia en las publicaciones científicas.

Una forma de evaluar la **confiabilidad** de una medición es hacerla dos veces con el mismo grupo de personas, y la correlación entre esas dos pruebas se denomina **confiabilidad por prueba-reprueba**. Sin embargo, este método comúnmente no resulta práctico o apropiado. Por ejemplo, el método no sería aplicable si, al realizar la prueba una vez, influye en la realización de la prueba por segunda vez (como sería el caso de una prueba de inteligencia).

Con muchas mediciones, tales como la mayoría de los cuestionarios, también se puede evaluar la **confiabilidad**, correlacionando el valor promedio de una mitad de los ítems con el valor promedio de la otra mitad. Por ejemplo, se podría correlacionar el valor de todos los ítems impares con el valor de todos los ítems pares. Si la persona está respondiendo coherentemente, deberíamos obtener una correlación alta. Este procedimiento se denomina **confiabilidad por división en mitades**

El problema que surge al utilizar el método de mitades es el modo en que se las divide. En muchos casos tiene sentido dividir los ítems en pares e impares, pero podría ocurrir que por casualidad esta división diera una correlación demasiado baja o demasiado alta. Afortunadamente, existe una solución más general: se puede dividir la prueba en mitades, de todas las formas posibles, y calcular la correlación utilizando cada una de las divisiones. El promedio de esas correlaciones se denomina **alfa de Cronbach** (α). (Existe una fórmula no demasiado compleja para realizar ese procedimiento, que produce el mismo resultado que promediar todas las posibles correlaciones entre mitades. Por supuesto, en la actualidad, alfa casi siempre se calcula con una computadora).

El alfa de Cronbach es la medida de **confiabilidad** más ampliamente utilizada, y también se la puede considerar como la descripción del grado en que cada ítem está asociado con cada uno de los otros ítems. Describe la coherencia general de la prueba, es decir, en qué medida las respuestas altas coinciden con las altas y las bajas con las bajas en todos los ítems de la prueba.

Generalmente, en psicología una prueba debería presentar una **confiabilidad** (medida a través del alfa de Cronbach) de al menos 0,7, y preferentemente cercana a 0,9, para que la prueba sea útil. Sin embargo, algunas veces se consideran adecuadas alfas de 0,6 ó menores.

Un contexto en el cual la **confiabilidad** es casi siempre discutida es en el de las publicaciones científicas, cuyo objetivo es, principalmente, la creación de una nueva medida. Por ejemplo, Sellers y sus colegas (1997) desarrollaron un cuestionario para evaluar la identidad de la raza negra entre americanos africanos. Al desarrollar la escala identificaron una cantidad de aspectos de la identidad de la raza negra, creando un MIBI (*Multidimensional Inventory of Black Identity*, Inventario multidimensional de identidad de la raza negra) que incluye diversas sub-escalas. Uno de los distintos métodos que utilizaron para evaluar la solidez de la escala como medida fue determinar la **confiabilidad** de cada sub-escala y, también, hacerlo con alumnos africanos americanos tanto en una universidad para alumnos de raza negra como en otra universidad en la que predominaban los alumnos de raza blanca. La tabla 17-5 indica los resultados de ese aspecto del estudio. (La línea correspondiente al "interés por lo público" está en blanco en la tabla, porque era una sub-escala que incluyeron originalmente pero luego descartaron durante el proceso de desarrollo de la medida). Sellers et al. resumen del siguiente modo los descubrimientos que presenta la tabla: "Las versiones revisadas de las escalas y sub-escalas del MIBI mostraron una adecuada coherencia interna [...] Los alfas de Cronbach de las sub-escalas iban desde un 0,60 bajo (aspecto privado) a un 0,79 (nacionalismo). Los alfas eran similares en cada una de las facultades" (p. 810).

Tabla 17-5.
Estadística descriptiva para el MIBI por facultad y para la muestra completa.

	Muestra completa			Universidad con predominio de raza blanca			Universidad afro-americana		
Escala	*α de Cronbach*	*M*	*SD*	*α de Cronbach*	*M*	*SD*	*α de Cronbach*	*M*	*SD*
Posición central	0,77	5,23	1,08	0,78	5,20	1,14	0,75	5,28	0,98
Interés priv.	0,60	6,25[a]	0,70	0,55	6,38	0,59	0,61	6,05	0,81
Interés púb.	—	—	—	—	—	—	—	—	—
Integración cultural	0,73	4,92[a]	0,91	0,66	5,16	0,80	0,74	4,55	0,94
Humanista	0,70	5,15[a]	0,84	0,68	5,33	0,80	0,69	4,87	0,81
Minoría	0,76	4,78[a]	0,82	0,75	4,82	0,80	0,77	4,70	0,86
Nacionalismo	0,79	4,27[a]	0,99	0,78	4,02	0,96	0,74	4,67	0,90

Nota: Priv = Privado; Pub = Público.

[a] Denota medias significativamente diferentes entre las dos muestras con un $\alpha = 0{,}01$.

Fuente: Sellers, R. M., Rowley, S. A. J., Chavous, T. M., Shelton, J. N., & Smith, M. A. (1997), tab. 2. "Inventario multidimensional de identidad de la raza negra: una investigación preliminar de la confiabilidad y validez del constructo". *Revista científica de psicología social y de la personalidad [Journal of Personality and Social Psychology]*, 73, 805-815. Copyright, 1997, por la Asociación Americana de Psicología. Reimpreso con autorización.

ANÁLISIS FACTORIAL

El **análisis factorial** se utiliza cuando el investigador ha medido a los participantes con respecto a una gran cantidad de variables. El análisis factorial indica al investigador qué variables tienden a agruparse, es decir, qué variables tienden a correlacionarse entre sí y no con otras. Cada agrupación de ese tipo (grupo de variables) se denomina **factor**. La conexión relativa de cada una de las variables originales con un factor es la **carga factorial** de esa variable en ese factor. (Las variables presentan cargas en todos los factores, pero generalmente tendrán cargas altas sólo en uno). Las cargas factoriales pueden considerarse como la correlación de la variable con el factor y, al igual que las correlaciones, van desde –1, asociación negativa perfecta con el factor, pasando por 0, ausencia de relación con el factor, hasta +1, correlación positiva perfecta con el factor. Normalmente, se considera que una variable contribuye significativamente en un factor sólo si presenta aproximadamente una carga de 0,3 ó mayor (ó de –0,3 ó menor). Algunos investigadores utilizan los niveles 0,35, 0,40, e incluso niveles más altos, como norma para decidir si una carga factorial es lo suficientemente importante como para considerar que la variable forma parte del factor.

El análisis factorial en sí mismo incluye una serie de fórmulas relativamente complejas que comienzan con las correlaciones entre todas las variables y terminan con una serie de cargas factoriales, así como también otros datos, tales como la cantidad de varianza, del total de variación entre las variables, que son explicadas por cada factor. En realidad existen varios métodos, de algún modo diferentes, para realizar un análisis factorial; así, el investigador cuenta con cierta li-

Tabla 17-6.
Cargas factoriales de ítems de las cuatro sub-escalas ideológicas.

Ítem	Factor 1	Factor 2	Factor 3	Factor 4
Integración cultural 5	**0,74**	0,43		
Integración cultural 4	**0,72**	0,35		
Integración cultural 6	**0,57**	0,41		
Integración cultural 7	**0,42**	0,63		
Integración cultural 2	**0,37**			
Integración cultural 3	**0,36**			–0,30
Integración cultural 1	**0,33**	0,48		
Integración cultural 9	**0,32**			
Integración cultural 8	**0,30**	0,41		–0,50
Humanista 7	0,41	**0,65**		
Humanista 1	0,31	**0,62**		
Humanista 6	0,36	**0,57**		
Humanista 2		**0,50**		–0,32
Humanista 8		**0,42**		
Humanista 4		**0,38**		–0,53
Humanista 5		**0,37**		–0,40
Humanista 3		**0,31**		–0,52
Humanista 9		**0,22**		
Minoría 8			**0,72**	
Minoría 3			**0,60**	
Minoría 9			**0,58**	
Minoría 5			**0,50**	
Minoría 1			**0,42**	
Minoría 4		0,34	**0,42**	
Minoría 7		0,31	**0,40**	
Minoría 2			**0,38**	
Minoría 6		0,35	**0,33**	
Nacionalista 7				**0,70**
Nacionalista 6				**0,63**
Nacionalista 1				**0,62**
Nacionalista 3		–0,40	0,30	**0,54**
Nacionalista 4		–0,33		**0,50**
Nacionalista 5		–0,54		**0,45**
Nacionalista 2		–0,51		**0,40**
Nacionalista 9		–0,50		**0,32**
Nacionalista 8		–0,51		**0,28**

Nota: sólo se enumeran las cargas mayores a 0,30, con excepción de los ítems Humanista 9 y Nacionalista 8. Los valores en negrita son los predichos por el MIBI.

Fuente: Sellers, R. M., et al. (1997), tab. 1. "Inventario multidimensional de identidad de la raza negra. Investigación preliminar de confiabilidad y validez de constructos". *Revista científica de psicología social y de la personalidad [Journal of Personality and Social Psychology]*, 73, 805–815. Copyright, 1997, por la Asociación Americana de Psicología. Reimpreso con autorización.

bertad para seleccionar entre una variedad de métodos, cada uno de los cuales puede dar resultados levemente diferentes.

Sin embargo, la parte más subjetiva del análisis factorial reside en el nombre que se le da al factor. Al leer una publicación científica que informa acerca de un análisis factorial, primero debería analizarse si el nombre que el investigador da a un factor describe adecuadamente las variables que lo conforman.

En el caso del estudio que acabamos de analizar, Sellers et al. también realizaron varios análisis factoriales como parte del desarrollo de su medida de identidad de la raza negra. La tabla 17-6 indica resultados de un análisis factorial de los ítems de sus cuatro escalas ideológicas. Los investigadores describen el análisis de la siguiente forma:

> La tabla [17-6] presenta la solución con cuatro factores del análisis factorial de las sub-escalas ideológicas. Debido al modo en el que se realizaron las operaciones con la escala ideológica, planteamos la hipótesis de que cada sub-escala ideológica se agruparía como un factor único, pero que podría haber cierta superposición en la solución final y las cargas serían moderadas [...] Para ser coherentes, y como método para reducir los ítems, se conservaron las nueve cargas superiores de cada sub-escala. Todas las cargas resultantes, excepto dos, fueron superiores a 0,30, ubicándose la mayoría en un rango moderado (de 0,40 a 0,65). En muchos casos, los ítems presentaban cargas en dos factores, pero la serie de factores de la solución final tenía cargas adecuadas para cada uno de los ítems de la sub-escala. Los cuatro factores explicaban aproximadamente el 56% de la varianza. En unos pocos casos, ítems que tenían cargas adecuadas en los factores de forma coherente con nuestro modelo presentaban, en efecto, una carga más alta en otro factor (p. ej. integración cultural 7, integración cultural 8). El análisis de los contenidos del ítem sugiere que esos ítems representan actitudes políticas coherentes con nuestra teoría acerca de las dos ideologías y, probablemente, presenten cargas altas en ambos factores en estudios subsiguientes (pp. 809-810).

MODELO CAUSAL

En el caso de las técnicas de modelo causal, al igual que en el análisis factorial, el investigador ha probado a una cantidad de personas según una cantidad de variables, pero a diferencia del análisis factorial, el objetivo de las técnicas de modelo causal es analizar si el patrón de correlaciones entre las variables se ajusta a la teoría previa del investigador con respecto a qué variables son la causa de cuáles otras.

Las técnicas de modelo causal son ampliamente utilizadas en psicología. Primero presentaremos el método antiguo del análisis de senderos y, después, pasaremos al método más moderno y más elaborado de modelo de ecuación estructural.

Análisis de senderos

En el **análisis de senderos**, el investigador crea un diagrama con flechas que conectan las variables. Las flechas o senderos indican las conexiones causa-efecto entre las variables según la teoría del investigador. Después, el investigador calcula coeficientes de senderos para cada uno de los senderos. El **coeficiente de senderos** es similar a beta en la regresión múltiple: indica en qué medida un cambio en la variable al comienzo de la flecha se relaciona con un cambio en la variable al final de la flecha. (El coeficiente se calcula de forma tal que excluye la influencia de cualquier otra variable que tenga flechas hacia la variable ubicada al final de la misma flecha).

Analicemos el siguiente ejemplo: MacKinnon-Lewis y sus colegas (1997) realizaron un estudio examinando las variables de predicción de la aceptación social, por parte de sus pares, de niños de 8 a 10 años de edad. Las principales variables de predicción que utilizaron fueron las calificaciones de los niños en cuanto a la aceptación o al rechazo de sus padres, las calificaciones de los pares en cuanto a aceptación y agresión, y los conflictos con hermanos según se observaron en una interacción experimental. Probaron varios modelos causales diferentes y llegaron a la conclusión de que el más apropiado era el que llamaron “modelo 1”.

La figura [17-1] representa gráficamente los coeficientes de senderos estandarizados del modelo 1, e indica que los hermanos en cuyas madres se percibía y observaba mayor rechazo, se reportaban y observaban más agresivos entre sí que aquellos hermanos cuyas madres mostraban menor rechazo. Más aún, los niños que experimentaban relaciones entre hermanos más agresivas tenían mayores probabilidades de que sus pares los consideraran agresivos y eran menos aceptados por ellos. Aunque no se reveló una influencia paterna directa en la agresividad entre hermanos, sí se evidenció un efecto indirecto como resultado del hecho de que una menor aceptación paterna estaba relacionada con un mayor rechazo por parte de la madre (p. 1027).

En el diagrama de senderos mencionado, los senderos más importantes presentaban coeficientes significativos en las direcciones predichas. Por lo tanto, MacKinnon-Lewis et al. interpretaron los resultados como un apoyo favorable para su teoría.

Modelo de ecuación estructural

El **modelo de ecuación estructural** lleva también el nombre, entre otros, de modelo de variable latente; otro nombre como por ejemplo **Lisrel**, es el nombre de uno de los programas de computación específicos utilizado para ese tipo de análisis. Básicamente, el modelo de ecuación estructural es sólo una extensión especial del análisis de senderos. Al igual que este último, incluye un diagrama de senderos con flechas entre las variables y coeficientes de senderos para cada flecha.

Sin embargo, el modelo de ecuación estructural presenta varias ventajas importantes con respecto al antiguo método de análisis de senderos. Una ventaja considerable es que el procedimiento nombrado proporciona, en primer lugar, un indicio general de la concordancia entre la teoría (según se describe en el diseño de senderos) y los datos. Esa indicación de concordancia general se denomina **índice de concordancia** o índice de "la bondad de ajuste". Se utilizan varios índices de concordancia diferentes pero, en general, se considera que una concordancia de 0,9 ó mayor es una concordancia adecuada (usualmente el máximo es 1).

En el modelo de ecuación estructural también se puede calcular una especie de prueba de significación, en cuanto a si los datos concuerdan con la teoría. Decimos una "especie de prueba de significación" porque la hipótesis nula, en este caso, establece que la teoría concuerda con los datos. Para decirlo en otras palabras, un resultado significativo implicaría que la teoría no concuerda con los datos; en otras palabras, ¡un investigador que intenta demostrar una teoría esperará que la prueba de significación arroje un resultado no significativo! Sin embargo, en muchos casos los participantes son tantos que, aun existiendo una concordancia adecuada, el resultado es significativo debido a que la potencia es tan alta que, incluso, un leve desvío de la concordancia adecuada resulta significativo; por lo tanto, algunas veces los investigadores informan una concordancia significativamente inadecuada pero luego la ignoran y se concentran en los índices de concordancia.

Una segunda ventaja considerable del modelo de ecuación estructural, con respecto al análisis de senderos común, es que el primero utiliza lo que denominamos variables latentes. Una **variable latente** es aquella que no es medida realmente sino que representa la variable real que el investigador desearía medir, pero sólo puede aproximar a través de medidas reales. Por ejemplo, una variable latente podría ser la clase social, la que el investigador intenta aproximar a través de distintas variables medidas, tales como el nivel de ingresos, los años de educación, el prestigio del empleo y los metros cuadrados de la vivienda. Ninguna de esas variables medidas puede sustituir adecuadamente la clase social (aunque algunas pueden hacerlo mejor que otras). Lo que se necesita es algún tipo de promedio ponderado el cual tenga también en cuenta que, en su conjunto, el grupo de variables medidas tampoco refleja en forma perfecta la variable latente.

En el modelo de ecuación estructural, los cálculos matemáticos se establecen de modo que la variable latente resulte una combinación de las variables medidas, combinadas de modo tal de utilizar

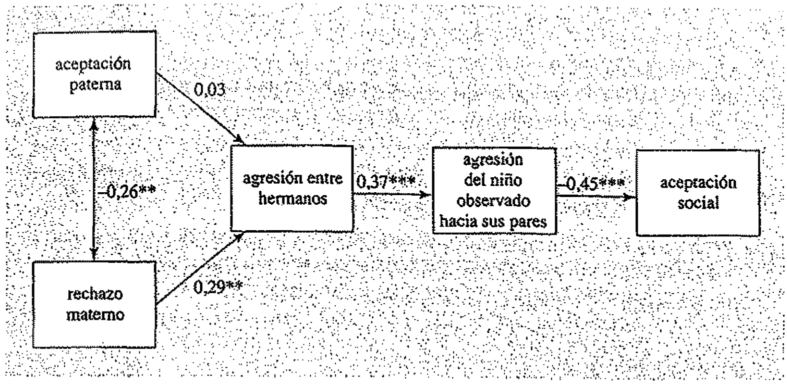


Figura 17-1.
Modelo de senderos de asociaciones entre variables de paternidad y maternidad, agresión entre hermanos, agresión entre pares y aceptación social. Se indican los coeficientes de senderos estandarizados.
$**p < 0{,}01$; $***\ p < 0{,}001$.

[**Fuente:** MacKinnon-Lewis, C., Starnes, R., Volling, B., & Johnson, S. (1997), fig. 1. "Percepciones de la paternidad y maternidad como variables de predicción de las relaciones entre los niños y sus hermanos o pares". *Psicología de Desarrollo [Developmental Psychology]* 33, 1024-1031. Copyright, 1997, por la Asociación Americana de Psicología. Reimpreso con autorización].

sólo lo que tienen en común entre sí. La idea es que lo que tienen en común es el verdadero valor con respecto a aquéllo de lo cual todas reflejan una parte. (Una variable latente es, en realidad, similar a un factor en el análisis factorial, en el sentido de que el factor no se mide directamente sino que representa una combinación ponderada de las diferentes variables que lo componen).

Tal como lo indica el ejemplo de la figura 17-2, en el diagrama de senderos de un modelo de ecuación estructural las variables que realmente se miden por lo general se representan en cuadrados o rectángulos, y las variables latentes en círculos u óvalos. Cabe destacar que en la figura las flechas van desde las variables latentes (las que se encuentran dentro de los círculos) hacia las variables medidas (aquellas dentro de los recuadros), para reflejar la idea de que la variable latente es la causa implícita de las variables medidas, siendo estas últimas la mejor forma posible de medir la verdadera variable latente.

También es importante observar que todas las otras flechas conectan variables latentes. En la mayoría de los casos, el modelo de ecuación estructural funciona de la siguiente manera: las variables medidas se utilizan para suplir las variables latentes, y el análisis se concentra en las relaciones causales (los senderos) entre estas últimas. (Finalmente, con respecto a las pequeñas flechas, que parecen no provenir de ningún lado, diremos que reflejan la existencia de cierto error (otras causas que no fueron medidas) que también afecta la variable. Son flechas de "error" o "alteración" que generalmente se omiten en las publicaciones científicas para que la figura resulte más simple, pero que de todos modos están implícitas).

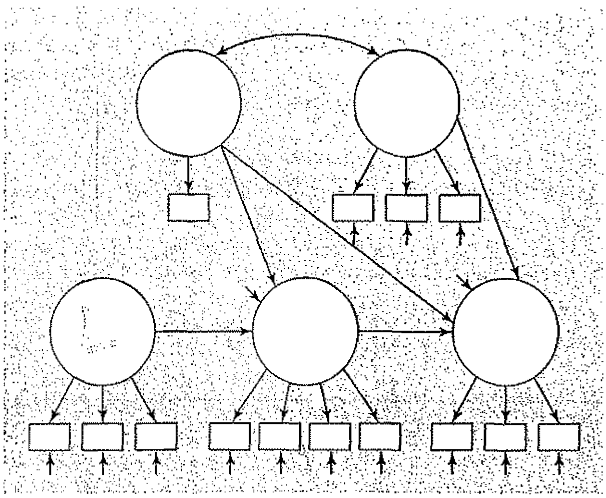


Figura 17-2.
Diagrama de senderos de un modelo de ecuación estructural

Ejemplo de modelo de ecuación estructural

El ejemplo que aquí brindamos proviene de un estudio realizado por Kwan y sus colegas (1997), el cual analizaba las variables de predicción de la satisfacción con respecto a la vida en general, concentrándose en el papel de la autoestima y la armonía social. Un dato particularmente interesante con respecto al estudio es que la predicción indicaba que la relación entre la autoestima y la armonía social con la satisfacción, en cuanto a la vida en general, sería diferente en las distintas culturas. En culturas más comunitarias, como es el caso de muchas culturas asiáticas, la armonía social tendría mayor importancia. Sin embargo, en culturas más individualistas, tales como la mayoría de las culturas americanas y europeas, la autoestima tendría mayor relevancia. Como parte del enfoque hacia las diferencias culturales, los investigadores también midieron el autoconcepto independiente (en qué medida una persona destaca el desarrollo y los logros personales) y el autoconcepto interdependiente (hasta qué punto una persona destaca el hecho de tener buenas relaciones y congeniar con los demás). El estudio se realizó con 389 alumnos universitarios de Estados Unidos y Hong Kong.

La figura 17-3 representa gráficamente los resultados básicos, concentrándose en los senderos entre las variables latentes. (Para que los diagramas sean simples, a veces las variables medidas ni siquiera se incluyen en el diagrama publicado). En este ejemplo en particular, los investigadores presentan dos series de coeficientes de senderos estandarizados para cada sendero. Los coeficientes de sendero que no están entre paréntesis corresponden a la muestra de los alumnos de Hong Kong; los que están entre paréntesis, a la muestra de Estados Unidos. Cabe destacar que el impacto del autoconcepto no es muy diferente en las dos culturas. Por ejemplo, los participantes de las dos culturas muestran aproximadamente el mismo grado de asociación entre la autoestima con la medida en la que una persona destaca la independencia. Lo más interesante, sin embargo, es que el sendero desde la autoestima a la satisfacción con la vida en general es mayor en el caso de la muestra de Estados Unidos, mientras que el sendero desde la armonía en las relaciones hacia la satisfacción con la vida es mayor en la muestra de Hong Kong.

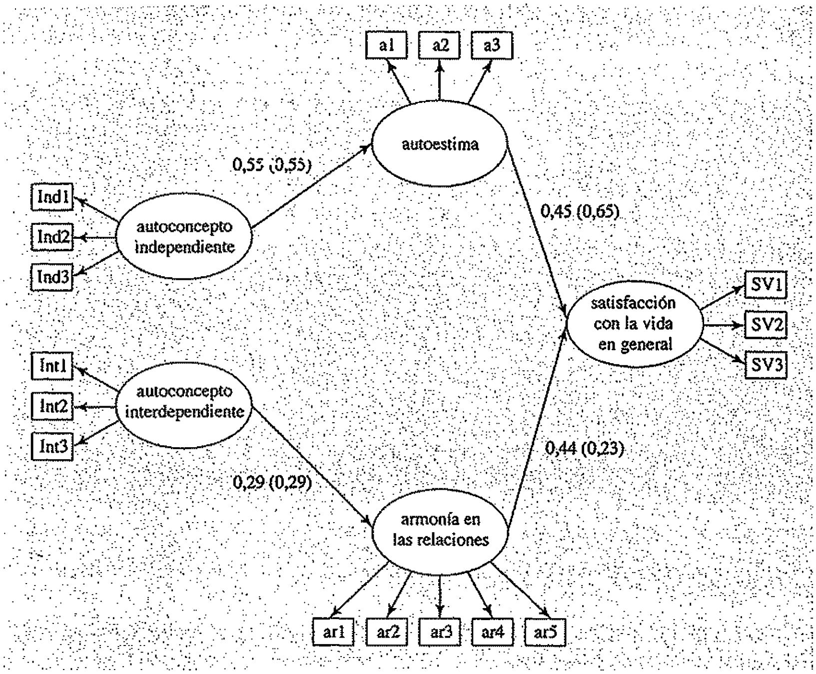


Figura 17-3.
Modelo final de la escala de autoconcepto. $N = 194$, en el caso de la muestra de Hong Kong y $N = 184$, en el de la muestra de Estados Unidos. Las elipses representan constructos latentes; los rectangulitos representan indicadores; las flechas que van desde los constructos latentes hacia los indicadores describen cargas factoriales, y las flechas que relacionan constructos latentes entre sí representan coeficientes de senderos. Se indican los coeficientes de senderos estandarizados y se omitieron las cargas factoriales y los errores de medición para que el diagrama resulte más claro. Los números entre paréntesis son los coeficientes correspondientes a la muestra de EE.UU., y los números que no están entre paréntesis son los coeficientes de la muestra de Hong Kong. Todos estos coeficientes resultaron significativos a $p < 0{,}05$ ó menor. [**Fuente:** Kwan, V. S., Bond, M. H., & Singelis, T. M. (1997), fig. 1. "Explicaciones panculturales de satisfacción con respecto a la vida en general: agregando la armonía en las relaciones a la autoestima". *Revista Científica de Psicología Social y de la Personalidad (Journal of Personality and Social Psychology)*, 73, 1038-1051. Copyright, 1997, por la Asociación Americana de Psicología. Reimpreso con autorización.]

Algunas limitaciones del modelo causal

Es importante ser consciente de que no existe demasiada magia detrás de estos métodos maravillosos, ya que los mismos siguen dependiendo siempre del razonamiento profundo del investigador. Supongamos que todos los senderos predichos de un diagrama de análisis de senderos resultan significativos, y que el modelo de ecuación estructural en general presenta una concordancia excelente. Aun en ese caso, es bastante probable que otros patrones de causalidad (otras formas de disponer las flechas) pudieran funcionar tan adecuadamente o mejor.

Otras alternativas podrían disponer las flechas en las direcciones contrarias o realizando conexiones diferentes, o bien, el patrón podría incluir variables adicionales que no aparecen en el diagrama original. Por lo tanto, todo tipo de modelo causal indica, en el mejor de los casos, que los datos son coherentes con la teoría, pero los mismos datos podrían también ser coherentes con teorías bastante diferentes. Lo ideal sería que el investigador pruebe diseños alternativos y pueda demostrar que los datos no concuerdan adecuadamente con ellos, pero que, al mismo tiempo, siempre pueden existir alternativas que el investigador ni siquiera haya pensado.

Además, el modelo causal, y todas aquellas técnicas que hemos estudiado hasta ahora y que dependen básicamente de las correlaciones, están sujetas a las precauciones que señalamos en los capítulos 3 y 4. La más importante de esas precauciones es la que acabamos de recalcar: la asociación no demuestra dirección de causalidad; es más, estas técnicas sólo tienen en cuenta en forma directa las relaciones lineales. Finalmente, si existe alguna restricción del recorrido, los resultados se distorsionan (generalmente tienden a arrojar menores coeficientes de senderos).

Por lo tanto, no debemos dejarnos arrollar por la sofisticación matemática de una técnica tal como el modelo de variable latente. Es verdad que la técnica resulta útil, a veces hasta maravillosamente útil, pero también es cierto que, si no se ha realizado una asignación aleatoria a los grupos, la dirección causa y efecto continúa siendo ambigua. Si las relaciones implícitas son curvilíneas o existen otras limitaciones, tales como la restricción del recorrido, por lo general es incluso más probable obtener resultados engañosos con los procedimientos más sofisticados que con las simples correlaciones bivariadas.

ANÁLISIS DE COVARIANZA

Hasta este punto del capítulo hemos analizado procedimientos estadísticos que hacen hincapié en las asociaciones entre variables, los cuales son básicamente elaboraciones sofisticadas de la correlación y la regresión. Ahora nos dedicaremos a los procedimientos que se basan en las diferencias entre las medias grupales, y que son esencialmente elaboraciones del análisis de varianza.

Entre los análisis mencionados anteriormente, una de las elaboraciones más ampliamente utilizadas es el ANCOVA. En este análisis, el investigador realiza un análisis de varianza común, pero antes ajusta las variables de modo de librarse del efecto de algunas variables adicionales no deseadas. Es decir, el ANCOVA es al análisis de varianza lo que la correlación parcial es a la correlación ordinaria. La variable controlada o excluida se denomina **covariable**. El resto de los resultados se interpretan como cualquier otro análisis de varianza.

Analicemos un ejemplo. Capaldi y Patterson (1991) realizaron un estudio acerca de la adaptación de niños al colegio primario, comparando la adaptación de niños que, desde su nacimiento, habían experimentado diferentes niveles de "transiciones paternas". Los diferentes niveles de transición paterna eran los siguientes: ausencia de transición, pérdida del padre, nuevo padrastro y dos o más padrastros nuevos. Los autores informan, "un ANOVA mostró que existían diferencias significativas entre los grupos de transición, $F(3, 170) = 7{,}53$, $p < 0{,}001$". (El patrón formado por las medias de los cuatro niveles coincidía con lo predicho en cuanto a que, a mayores transiciones paternas, más insatisfactoria era la adaptación del niño).

Sin embargo, los investigadores eran conscientes de que las familias de los niños que formaban los cuatro niveles de transición pertenecían a diferentes SSE (situaciones socio-económicas) y tenían diferentes niveles de ingreso. ¿Podrían estas diferencias, y no las diferencias en cuanto a niveles de transición, ser la causa implícita de las diferencias de adaptación?

> Después, probamos la hipótesis de que las diferencias entre los grupos de transición eran fundamentalmente una función de las diferencias de SSE e ingresos. Para probar esa presunción se realizó un ANOVA con las covariables de SSE e ingreso per cápita. La diferencia entre los grupos de transición continuó siendo significativa, $F(5, 167) = 4{,}0$, $p < 0{,}01$ (pp. 492-493).

(El patrón de medias fue el mismo en este análisis que en el original). Aunque ellos no utilizaron el término específico, un ANOVA con covariables es un análisis de covarianza.

ANÁLISIS DE VARIANZA MULTIVARIADO Y ANÁLISIS DE COVARIANZA MULTIVARIADO

Todos los procedimientos que hemos tratado hasta aquí en el libro, incluso los tratados en este capítulo, incluyen sólo una variable dependiente. Pueden existir dos o incluso muchas variables independientes o de predicción, como en el caso de la regresión múltiple o el análisis factorial de varianza; pero en todos los casos, sólo ha existido una variable dependiente.

En esta sección, analizamos temas de estadística multivariada, es decir, procedimientos utilizados cuando existen dos o más variables dependientes. Específicamente, nos concentramos en los dos procedimientos multivariados más ampliamente utilizados: elaboraciones multivariadas del análisis de varianza y del análisis de covarianza. Son versiones del análisis de varianza y covarianza que pueden manejar más de una variable dependiente.

El MANOVA es un análisis de varianza en el que pueden existir varias variables dependientes. Usualmente, estas variables dependientes son diferentes medidas de prácticamente la misma cosa, como por ejemplo, tres escalas diferentes de compromiso político o tres pruebas diferentes de habilidad para la lectura. Los resultados del MANOVA se interpretan básicamente del mismo modo que los de un análisis de varianza común. Supongamos que se estudian tres grupos y se mide a cada participante según cuatro variables dependientes. El MANOVA daría un F general, y un nivel de significación de la diferencia entre los tres grupos, en términos del grado de desigualdad en cuanto a la combinación de las cuatro variables.

Cuando los investigadores encuentran una diferencia significativa general entre grupos a través de un MANOVA, se entiende que los grupos difieren en la combinación de las variables dependientes. Comúnmente, los investigadores también querrán saber si los grupos se diferencian en alguna o en todas las variables dependientes tomadas en forma individual. Por lo tanto, es común que después de un MANOVA se realice una serie de análisis de varianza comunes, uno para cada una de las variables dependientes. Los análisis de varianza individuales se denominan a veces análisis de varianza "univariado" (en contraposición con el análisis multivariado"), porque cada uno tiene sólo una variable dependiente. Al igual que con otros análisis univariados, generalmente incluyen comparaciones múltiples, tales como las pruebas de contrastes lineales.

Analicemos un ejemplo. DeGarmo y Forgatch (1997) analizaron un grupo de madres divorciadas, concentrándose en el respaldo que recibían de su confidente más cercano. El confidente era a veces un amigo íntimo, a veces un miembro de la familia, y otras una pareja con la que convivían. En el estudio, tanto las madres como los confidentes fueron entrevistados conforme a varias medidas; también se los observó interactuando, y los investigadores codificaron la interacción en forma sistemática. Los distintos métodos utilizados crearon unas cuantas medidas de la relación entre la madre y su confidente íntimo, incluyendo tres medidas de respaldo por parte del confidente, cuatro medidas de aspectos negativos del confidente y cuatro medidas de la intimidad de la relación.

Un aspecto del estudio se concentraba en cómo difería la relación con el confidente en el caso de que el mismo fuera un amigo, un miembro de la familia o la pareja con la que convivían. DeGarmo y Forgatch describieron el análisis de la siguiente manera:

> Se realizaron análisis de varianza multivariado y univariado con los indicadores de respaldo de aspectos negativos y de intimidad con respecto a amigos cercanos, familiares y parejas de convivencia. La tabla [17-7] indica los valores medios, las pruebas de diferencias y los contrastes significativos.
> Se encontraron diferencias significativas entre los tipos de relación MANOVA de los indicadores, $F(20, 254) = 4{,}10, p < 0{,}001$ (p. 340).

Tabla 17-7.
Medias y desvíos estándar de los indicadores de constructo según el tipo de relación con el confidente.

	Amigo (1)		Familiar (2)		Pareja (3)			
Indicadores de constructo	*M*	*SD*	*M*	*SD*	*M*	*SD*	*F(2, 135)*	*Contrastes significativos*
Respaldo observado en el confidente								
Interpersonal	3,34	0,67	3,35	0,63	2,92	0,65	5,93**	1, 2 > 3
Afabilidad	3,39	0,86	3,24	0,94	2,68	1,21	6,58**	1, 2 > 3
Emocional	1,04	0,36	0,96	0,37	0,69	0,35	12,17***	1, 2 > 3
Aspectos negativos del confidente								
Irritabilidad según el propio informe	1,91	0,84	1,70	0,70	2,25	0,65	5,27**	3 > 2
Irritabilidad según el informe de intimidad	1,36	0,50	1,33	0,35	1,48	0,40	1,65	
Depresión	1,06	0,32	0,93	0,36	0,95	0,34	2,02	
Intimidad de la relación								
Intimidad según el informe de la madre	3,18	0,73	3,19	0,75	3,65	0,58	5,94**	3 > 1, 2
Intimidad según el informe del confidente	3,05	0,78	3,29	0,69	3,48	0,64	4,62**	3 > 1
Complejidad según informe de la madre	1,91	0,84	2,29	0,74	2,87	0,33	22,52***	3 > 1, 2
Complejidad según informe del confidente	2,01	0,74	2,19	0,75	2,73	0,55	13,36***	3 > 1, 2

Nota: *ns* = 65, 33 y 40 para cada tipo de relación, es decir, amigos, familiares y parejas respectivamente.
** $p < 0{,}01$; *** $p < 0{,}001$.

Fuente: DeGarmo, D. S., & Forgatch, M. S. (1997), tab. 2. "Determinantes del respaldo observado en el confidente hacia las madres divorciadas": *Revista Científica de Psicología Social y de la Personalidad [Journal of Personality and Social Psychology]*, 72, 336-345. Copyright, 1997, por la Asociación Americana de Psicología. Reimpreso con autorización.

DeGarmo y Forgatch debatieron después los resultados del análisis de varianza univariado y los contrastes subsiguientes. Por ejemplo, observaron que el "análisis de varianza mostró un patrón según el cual se observaba que las parejas proporcionaban menor respaldo" (p. 340).

Un análisis de covarianza en el que existe más de una variable dependiente se denomina MANCOVA. La diferencia entre un MANCOVA y un análisis de covarianza común es, precisamente, paralela a la diferencia entre un MANOVA y un análisis de varianza común. Es decir, un MANCOVA es un MANOVA en el que existe una o más covariables (variables que se mantienen constantes o son controladas).

REPASO DE LAS DISTINTAS TÉCNICAS ESTADÍSTICAS

La tabla 17-8 indica de modo sistemático las diferentes técnicas que hemos estudiado en este capítulo, junto con los otros procedimientos paramétricos tratados en el resto del libro. Se puede poner a prueba lo aprendido cubriendo la columna de la derecha y adivinando de qué procedimiento estadístico se trata.

Tabla 17-8.
Principales técnicas estadísticas.

Asociación o diferencia	Cantidad de Variables independientes	Cantidad de Variables dependientes	¿ Se controla alguna variable?	Nombre de la técnica
Asociación	1	1	No	Correlación/regresión bivaribles
Asociación	Cualquiera	1	No	Regresión múltiple (incluso la regresión jerárquica y gradual)
Asociación	1	1	Si	Correlación parcial
Asociación	Muchas, no diferenciadas		No	Coeficiente de confiabilidad Análisis factorial
Asociación	Muchas, con patrones causales especificados			Análisis de senderos Diseño de ecuación
Diferencia	1	1	No	Análisis de covarianza en sentido único; prueba *t*
Diferencia	Cualquiera	1	No	Análisis de covarianza
Diferencia	Cualquiera	1	Si	Análisis de covarianza
Diferencia	Cualquiera	Cualquiera	No	Análisis de covarianza multivariable
Diferencia	Cualquiera	Cualquiera	Si	Análisis de covarianza multivariable

CONTROVERSIA: ¿DEBERÍA SER CONTROVERTIDA LA ESTADÍSTICA?

La mayoría de los libros de estadística, incluso éste, enseñan métodos estadísticos de un modo bastante estereotipado, casi como impartiendo una verdad absoluta. Sin embargo, a medida que avanzamos, también hemos intentado mezclar esa prolija imagen con nuestras exposiciones acerca de las distintas controversias. Usualmente, se considera que esto resulta confuso para los alumnos. (Aunque cuando esos alumnos aprendieron otras áreas de la psicología, la comprensión de las mismas se construyó, eso esperamos, a partir de la presentación de controversias: la investigación de tal persona demostró tal cosa, pero el estudio realizado por esta otra persona mostró una imperfección en la investigación anterior, mientras que el estudiante de la primera demostró que era una excepción, y así sucesivamente). Por lo tanto, en esta última sección de controversias intentaremos crear aún más confusión.

En el cuadro 17-1 describimos el desarrollo histórico de la estadística actual, a partir de un híbrido de dos visiones diferentes conocidas como los métodos de Fisher y de Neyman-Pearson. Se suponía que ese matrimonio pondría fin a la lucha con respecto a cuál es el método más adecuado, pero de hecho, aunque la mayoría de los psicólogos se sienten conformes con el híbrido, otros, tales como Gigerenzer y sus socios (Gigerenzer & Murray, 1987; Gigerenzer et al. 1989; Sedlmeier & Gigerenzer, 1989), no están para nada satisfechos. Tampoco lo están Jacob Cohen (1990) ni Robert Rosenthal (p. ej. Rosnow & Rosenthal, 1989b), dos psicólogos muy conocidos por sus contribuciones a las técnicas estadísticas y cuyos trabajos sobre temas tales como potencia, tamaño de efecto, hipótesis nula, meta-análisis y otros hemos mencionado a lo largo de todo el libro.

Gigerenzer y Murray (1987) sostienen que los puntos de vista de Fisher y de Pearson y Neyman, los cuales para estos mismos estadísticos de la primera hora siempre fueron fundamentalmente contradictorios, han sido mal interpretados y erróneamente empleados al combinarlos. El matrimonio se realizó enteramente por conveniencia, pensando muy poco en los efectos a largo plazo. Gigerenzer y Murray consideran al híbrido como el resultado de que muchos de los primeros textos de estadística hayan sido escritos bajo la influencia del dógmatico y persuasivo Sir Ronald Fisher (cabe recordar el cuadro 11-1). Pero luego, después de la Segunda Guerra Mundial, se conoció la visión de Pearson-Neyman, y tuvo que ser integrada sin admitir que los textos originales podían haber estado equivocados. (El deseo era presentar la psicología como una ciencia, con base en un método de toma de decisiones unificado, mecánico y sin defectos).

El resultado de todo ese proceso, afirman Gigerenzer y Murray, es el abandono de la controversia y los métodos alternativos, al igual que textos sobre estadística "repletos de confusión conceptual, ambigüedad y errores" (p. 23). Más aún, ellos sostienen que estos métodos estadísticos dominantes, que originalmente fueron sólo herramientas, actualmente definen el modo en que los psicólogos consideran la propia percepción y cognición humanas (cabe recordar los cuadros 12-1 y 13-1).

En líneas más generales, los actuales e intensos debates sobre pruebas de significación que tratamos en los capítulos 6 a 8 son parte de esta tendencia mayor de reapertura de controversias por mucho tiempo olvidadas.

Por último, nos resta decir que la mayoría de los psicólogos y estadísticos se sienten bastante cómodos con los métodos transmitidos por los textos actuales. El tiempo, y un cuidadoso razonamiento, nos dirán si esa mayoría realmente debería sentirse tan satisfecha; pero nadie lo hará por nosotros sino que tendremos que hacerlo juntos. Por lo tanto, sinceramente deseamos que, una vez que el alumno domine los métodos contenidos en este libro, tenga la confianza suficiente para ir más allá y no se conforme con continuar aplicando dentro de veinte años estos mismos métodos de forma insensata y mecánica. Aquellos psicólogos que leen o realizan investigaciones, cualesquiera sean sus otros intereses, también deben ser buenos ciudadanos dentro de la disciplina en general. Deben acompañar, aunque sea un poco, los desarrollos en cuanto a los métodos de análisis de datos, aceptando, e incluso exigiendo, cambios cuando sean justificados. Después de todo, si nuestras herramientas se tornan anticuadas, ¿qué esperanza queda para nuestros descubrimientos?

Cuadro 17-1. La asociación forzada de Fisher y Pearson.

Demos una última mirada a la historia del desarrollo de métodos estadísticos en el área de la psicología. De esta manera será posible agregar algunos datos interesantes. Ya dijimos en el cuadro 11-1, que Sir Ronald Fisher prácticamente inventó el método experimental tal como se utiliza en la actualidad; que el método surgió de su trabajo en la agricultura (principalmente sobre la fertilidad del suelo, el peso de los cerdos y el efecto del abono en las plantaciones de papas); que era un hombre con quien resultaba difícil congeniar, y que Fisher y otro gran estadístico británico, Karl Pearson, eran enemigos.

Bien. Pearson tenía un hijo, Egon, quien trabajaba en el Laboratorio Galton, precedido por su padre, en la Facultad de la Universidad, en Londres. En 1925, el joven Egon formó una amistad perdurable con Jerzy Neyman, un joven catedrático de la Universidad de Varsovia que acababa de llegar al Laboratorio Galton. En los años siguientes, los dos trabajarían muy estrechamente.

En 1933, Karl Pearson se retiró. Irónicamente, Fisher recibió el antiguo puesto de Pearson como jefe del Departamento de Eugenesia, originalmente fundado por Galton. Como resultado de la enemistad entre Fisher y su colega de mayor edad, Pearson, se creó un nuevo Departamento de Estadística que estaría al mando de Egon, el hijo de Pearson, con el fin de acariciar las plumas del pájaro que se retiraba.

Por más firmeza que Pearson y su amigo Neyman hayan puesto para intentar evitar la continuación de la vieja enemistad entre Sir Ronald y su colega de mayor edad Pearson, pronto el enfrentamiento se tornó más punzante que nunca. En realidad, Pearson y Neyman estaban mucho más de acuerdo en muchos aspectos con las ideas de Fisher que con las de Karl Pearson, pero sus extensiones y elaboraciones de los métodos de Fisher, aunque pretendían ser cordiales, enfurecían al malhumorado Sir Ronald (después de todo, el alumno no querrá cambiar su especialización por historia, ¡recordar correctamente estos nombres es al menos tan difícil como lo fue aprender estadística!).

¿Qué se discutía? Para simplificar una serie de ideas muy complejas, diremos que Fisher había rechazado lo que se denomina la teoría bayesiana, un enfoque global sobre la estadística que hemos mencionado en el capítulo 5, el cual sostiene que la investigación científica se realiza para adaptar opiniones preexistentes en vista de las nuevas evidencias a medida que se recolectan. En desacuerdo, Fisher sostenía que la inferencia inductiva se realiza principalmente desaprobando objetivamente la hipótesis nula, y no probando probabilidades previas a las que se había arrivado subjetivamente.

Fisher era excepcionalmente dogmático con respecto a sus ideas, refiriéndose a su método como "absolutamente riguroso" y "perfectamente riguroso"; lo llamaba el único caso de "inferencia inequívoca"; tenía una gran mente y escribió muchísimo, haciéndose muy influyente en el mundo entero.

Pearson y Neyman también rechazaron la teoría bayesiana, pero propusieron el método de prueba de dos hipótesis opuestas en lugar de una sola hipótesis nula. Como resultado de esa innovación, habría dos tipos de errores: los errores Tipo I serían aquellos en los que la hipótesis nula se rechaza aun cuando es verdadera (y a la probabilidad de ese error la denominaron alfa o nivel de significación, ¿resulta familiar?) Los errores Tipo II serían aquellos en los que la hipótesis de investigación se rechaza aun cuando es verdadera (y la probabilidad de ese error era beta. ¿Esto también resulta familiar? El impacto de cada tipo de error, en el objetivo del investigador, indicaría cuál de ellos era preferible minimizar, ya que Neyman y Pearson pensaban con frecuencia en términos de investigación aplicada. Fisher nunca mencionó ninguna hipótesis excepto la nula y, por lo tanto, nunca tuvo en cuenta los errores Tipo II.

Ahora queda claro lo que sucedió: la estadística es un híbrido de las ideas de Fisher con las de Pearson y Neyman, las últimas agregadas cuando ya no pudieron ignorarse. El concepto de probar la hipótesis nula proviene de Fisher; los conceptos algo menos influyentes de error Tipo II, beta, potencia y tamaño del efecto, de sus enemigos más jóvenes.

Fue una asociación que probablemente ninguno de ellos hubiera aprobado, ya que, con el tiempo, ambos lados consideraron sus propios métodos fundamentalmente opuestos a los del otro. Fisher comparaba a Neyman y a Pearson con el estereotipo de los soviéticos de su tiempo, en cuanto a su determinación de reducir la ciencia a la tecnología "en el amplio esfuerzo organizado de un plan quinquenal para la nación" y, además, comentó sarcásticamente después de que Neyman finalizara su discurso frente a la Royal Statistical Society (Real Sociedad de Estadística) en Londres, que Neyman debería haber elegido un tema "sobre el cual pudiera hablar con autoridad". Neyman, por su parte, declaró que los métodos de prueba de Fisher eran "en un sentido matemáticamente especificable, peores que inútiles". ¡Ah, qué racional!

Si bien el debate actual acerca del rol de la prueba de significación en el área de la psicología (*véase* las secciones de "Controversias" de los capítulos 6-8) no es tan estridente, sí conserva algo de la resonancia de los viejos tiempos. Por ejemplo, dos de los principales contendientes (Schmidt & Hunter, 1997) comentan que "todas las objeciones" a los argumentos a favor de su posición "son lógicamente deficientes" (p. 38) y que, "aunque cada una de estas objeciones parece plausible e incluso convincente para muchos investigadores, en realidad son un fracaso lógica e intelectualmente" (pp. 61-62). En un artículo publicado casi al mismo tiempo, dos de los principales contendientes del lado opuesto (Cortina & Dunlap, 1997) describieron los argumentos del otro lado como "construidos sobre supuestos defectuosos, ejemplos engañosos y errores en cuanto a ciertos conceptos críticos" (p. 170). Los comentarios que hemos escuchado de ambos lados, en ambientes menos formales, han sido aún menos contenidos.

Como puede observarse, a través de las historias relatadas en los cuadros de este libro, la estadística es, para bien o para mal, producto del intelecto y de las pasiones humanas funcionando en forma conjunta (idealmente, por el bien de la ciencia, aunque la última en menor grado). Los resultados no siempre han sido perfectos, pero pueden resultar mucho más interesantes de lo que parecerían a primera vista.

CÓMO LEER RESULTADOS EN PUBLICACIONES CIENTÍFICAS QUE INCLUYEN TÉCNICAS ESTADÍSTICAS QUE NO NOS RESULTAN FAMILIARES

Sobre la base de lo aprendido en este capítulo y en todo el libro, el alumno debería estar bien preparado para leer y comprender, al menos en forma general, los resultados de la mayoría de las publicaciones científicas psicológicas. Sin embargo, de cuando en cuando se encontrará con nuevas técnicas (y a veces nombres no familiares para viejas técnicas). Le sucede incluso a investigadores experimentados. ¿Qué debemos hacer entonces cuando nos encontramos con elementos de los que nunca hemos escuchado hablar?

El primer paso es no desesperarse. En la mayoría de los casos puede deducirse la idea básica. Casi siempre se establecerá el nivel p y debería indicarse claramente el patrón de resultados que se considera significativo o no. Además, generalmente habrá algún indicio acerca del tamaño del efecto, del grado de asociación o de la magnitud de la diferencia. Si la técnica estadística se refiere a la asociación entre algunas variables, probablemente sea más fuerte a medida que el resultado se acerque a 1, y más débil a medida que el resultado se acerque a 0. En una situación de este tipo no debemos esperar comprender cada palabra, sino intentar captar lo que sea posible con respecto al significado del resultado.

Analicemos un ejemplo. Biernat y Wortman (1991) realizaron un estudio acerca de la vida hogareña de mujeres profesionales. Cerca del comienzo de la sección de resultados, los investigadores mencionan que, en algunos de sus análisis, compararán mujeres académicas con mujeres de negocios. Por lo tanto, explican, controlaron si las variables a comparar aparentemente cumplían el supuesto de iguales varianzas poblacionales. Con respecto a una variable, comentaron: "La variabilidad en la educación era mayor en el caso de las mujeres de negocios ($SD = 1,26$) que en el de las mujeres académicas ($SD = 0,12$), C de Cochran $(2, 136) = 0,99$, $p < 0,0001$". (p. 848)

Probablemente, el alumno que se encuentre con el informe anterior nunca haya escuchado hablar de la "C de Cochran". Sin embargo, por el contexto, puede imaginarse que se trata de una prueba de significación que compara la variabilidad de dos grupos. Probablemente no pueda calcular lo que significan exactamente las cifras entre paréntesis despúes de C de Cochran, o a que se refiere el 0,99, pero sí puede comprender el "$p < 0,0001$", que indica que la diferencia de variabilidad entre los dos grupos es significativa. Podría llegar aún más lejos y observar directamente los dos desvíos estandar, que dan una idea bastante clara de lo muy diferentes que son las variabilidades en los dos grupos.

Supongamos que el alumno realmente no pueda captar absolutamente nada de una técnica estadística utilizada en una publicación científica. En ese caso, puede intentar buscar el procedimiento en un libro de estadística. Los libros de estadística intermedia y avanzada a veces son una buena opción, aunque hay que ser conscientes de que intentar comprender un texto de nivel intermedio, sin ayuda, puede resultar difícil. Muchos de esos textos tienen una orientación fundamentalmente matemática, incluso los textos más accesible utilizarán cada uno sus propios símbolos; por lo tanto, puede resultar difícil comprender sus descripciones de un método en particular sin haber leído todo el libro. Una mejor solución, en este caso, tal vez sea pedir ayuda a un profesor o alumno graduado en el campo en cuestión. Si el alumno conoce los principios básicos aprendidos a través de este libro, estará preparado para comprender los principios fundamentales de las explicaciones que reciba.

Si el alumno se encuentra a menudo con técnicas estadísticas que no comprende, la mejor solución es asistir a otros cursos de estadística. El siguiente curso, en la mayoría de los programas en el área de la psicología, es un curso intermedio que se concentra principalmente en el análisis de varianza, y puede llegar a abarcar hasta cierto nivel de la regresión múltiple. Estos tipos de cursos serán particularmente útiles para los alumnos que tengan intenciones de realizar un posgrado en psicología, en donde la estadística será una herramienta crucial en todas las investigaciones que realicen. Cursos de ese tipo los ayudarán a prepararse para el posgrado. Además, un buen desempeño en ese tipo de cursos produce una impresión extremadamente buena en aquellos que evalúan las solicitudes de ingreso a los mejores programas para graduados. (También podemos decir que, según nuestra experiencia, lo más probable es que el alumno disfrute con los otros estudiantes que conozcan en esos cursos. Los alumnos que asisten a cursos intermedios de estadística aplicada a la psicología no son todos fenómenos de las estadísticas, pero casi siempre son alumnos muy motivados y brillantes que seguramente compartirán los objetivos del lector). De hecho, a algunas personas, la estadística le resulta tan fascinante ¡que deciden hacer de ella una carrera!

En líneas más generales, constantemente se están inventando nuevos métodos estadísticos. Todos los psicólogos encuentran en las publicaciones científicas que leen números y símbolos que no le son familiares; pero finalmente los resuelven del mismo modo que lo hará el lector. Y tenemos plena confianza en ello debido a que ha llegado ileso y bien preparado a las útlimas páginas de este libro. Ha dominado la introducción detallada de un tema complejo; por ello, debería confiar en que con un poco de tiempo y esmero será capaz de comprender cualquier otro tema de estadística más avanzado. Por eso queremos felicitar al lector por sus logros.

Resumen

En la regresión múltiple jerárquica, las variables de predicción se incluyen en la regla de predicción en forma planificada y secuencial, permitiendo al investigador determinar la contribución relativa de cada variable siguiente por encima de aquellas ya incluídas. La regresión múltiple por pasos es un procedimiento de exploración en el que se examinan las potenciales variables de predicción para encontrar la mejor variable de predicción; luego se examinan las variables restantes para encontrar la variable de predicción que, en combinación con la primera, produce la mejor predicción. El proceso continúa hasta que agregar la mejor variable restante no aporta ninguna mejora significativa.

La correlación parcial describe el grado de correlación entre dos variables a la vez, que mantiene constante otra u otras variables.

Los coeficientes de confiabilidad indican en qué medida las puntuaciones de una prueba son internamente coherentes (usualmente con el alfa de Cronbach) o coherentes a través del tiempo (confiabilidad por prueba y reprueba).

El análisis factorial identifica agrupaciones de variables que se correlacionan en el máximo grado posible entre sí, y en el mínimo grado posible con otras variables.

El análisis causal examina si las correlaciones entre diversas variables son coherentes con un patrón sistemático e hipotético de relaciones causales entre ellas. El análisis de senderos describe esas relaciones con flechas que van desde la causa al efecto, con un coeficiente de senderos para cada flecha que indica la influencia de la hipotética variable causal en la hipotética variable de efecto. El modelo de ecuación estructural es una versión avanzada del análisis de senderos, que incluye variables latentes teóricas que no son medidas (cada una de las cuales está formada por los elementos comunes de diversas variables medidas). El modelo también ofrece medidas de la concordancia general de los datos con el patrón causal hipotético.

El análisis de covarianza es un análisis de varianza que controla una o más variables. El análisis de varianza multivariado es un análisis de varianza con dos o más variables dependientes. El análisis de covarianza multivariado es un análisis de covarianza con dos o más variables dependientes.

En los últimos años, los psicólogos han comenzado a reexaminar los principios básicos de la estadística que utilizamos creando la posibilidad de controversia acerca de aquello que, con frecuencia, había sido considerado incontrovertible en el pasado.

En general, es posible captar la idea principal de un procedimiento estadístico no familiar teniendo presente que probablemente se refiere a asociaciones entre variables o diferencias entre grupos, que el valor *p* indica la significación de esa asociación o diferencia, y que probablemente el procedimiento incluya algunos números a partir de los cuales podamos tener una idea del grado de asociación o diferencia.

Términos clave

- ANCOVA.
- Controlar.
- Covariable.
- Alfa de Cronbach (α).
- Factor.
- Análisis factorial.
- Carga factorial.
- Índice de concordancia.
- Regresión múltiple jerárquica.
- Mantener constante.
- Variable latente.
- Lisrel.
- MANCOVA.
- MANOVA.
- Estadística multivariada.
- Correlación parcial.
- Coeficiente de correlación parcial.
- Excluir.
- Análisis de senderos.
- Coeficiente de senderos.
- Confiabilidad.
- Confiabilidad por división en mitades.
- Regresión múltiple gradual.
- Modelo de ecuación estructural.
- Confiabilidad por prueba y reprueba.

Ejercicios

En los ejercicios 1 al 5 de la serie I, y en los ejercicios 1 al 4 de la serie II, se espera que el alumno explique sólo el significado general de los resultados en la forma en que los diferentes métodos fueron descriptos a lo largo del capítulo. Por supuesto que no se espera que el alumno describa la lógica de los procedimientos estadísticos tratados aquí del mismo modo en el que lo ha estado haciendo en los capítulos anteriores.

En la última parte del libro se indican las respuestas a la serie I de ejercicios.

SERIE I

1. Parte de un estudio realizado por Lindzey et al. (1997) examinaba de qué modo la reciprocidad en la interacción entre padre e hijo predecía la capacidad de integrarse a la vida social de niños en edad preescolar. En el estudio, cada niño o niña era observado interactuando con su padre en una situación estandarizada. Las interacciones se clasificaban de forma tal que producían medidas sobre quién iniciaba las actividades de juego además de la reciprocidad (equilibrio) en el cumplimiento de la iniciativa de juego del otro. Los investigadores también pidieron a los maestros del niño que calificaran la capacidad de cada niño para integrarse a la vida social con los otros niños de la ecuela. Descubrieron correlaciones entre la reciprocidad padre-hijo y la capacidad del niño para integrarse a la vida social. Sin embargo, les preocupaba saber hasta qué punto la medida de reciprocidad podría estar mezclada con el grado en el que los niños y los padres tomaban la iniciativa individualmente.

Por ende, realizamos una serie de análisis de regresión jerárquica para analizar si el cumplimiento recíproco de padre e hijo [...] aportaba contribuciones únicas a la predicción de la capacidad del niño para adaptarse a la vida social después de tener en cuenta el comportamiento de cada individuo [...] Los índices de iniciativa del padre y del niño fueron ingresados en primer lugar y justificaban el 3% de la varianza ($p = 0{,}57$). El cumplimiento recíproco de padre e hijo fue ingresado en segundo lugar y justificaba un significativo 18% adicional ($p = 0{,}01$) de la varianza de la calificación realizada por los maestros en cuanto a la capacidad de los niños para integrarse socialmente. (pp. 532-533).

Explique el método y el resultado a una persona que en general está familiarizada con la regresión múltiple común pero que nunca ha escuchado hablar de la regresión múltiple jerárquica.

2. Boyd y Gullone (1997) realizaron un estudio acerca de la angustia y la depresión con una muestra de 783 adolescentes que asistían a la escuela en Melbourne y sus alrededores, en Australia. Para medir la angustia utilizaron la RCMAS (*Revised Children's Manifest Anxiety Scale*, Versión revisada de la escala de angustia manifiesta en niños). Al tratar la medida en la sección Métodos, los investigadores realizaron la siguiente observación: "Las estimaciones de confiabilidad del coeficiente alfa, en cuanto a la coherencia interna de la RCMAS, iban de 0,42 a 0,87" (p. 192). Explique los resultados a alguien que está familiarizado con la correlación pero que nunca ha escuchado hablar de la confiabilidad o de los cálculos estadísticos relacionados con ella.

3. Fawzi et al. (1997) realizaron un estudio para evaluar si la manera usual de conceptuar el PTSD, tal como lo describe la cuarta edición del *Manual estadístico y de diagnóstico de trastornos mentales [Diagnostic and Statistical Manual of Mental Disorders]* (DSM-IV), se aplica a refugiados vietnamitas en los Estados Unidos. Como parte del estudio, se entrevistaron 74 refugiados (en su lengua nativa) con respecto a varios síntomas PTSD y a los hechos traumáticos que habían experimentado (p. ej. torturas). Como se esperaba, la cantidad de síntomas PTSD estaba correlacionada con la cantidad de hechos traumáticos. En un análisis más amplio del patrón de síntomas (qué síntomas se agrupan entre sí), realizaron un análisis factorial a través del cual obtuvieron cuatro factores.

De acuerdo con el DSM-IV, los primeros tres factores representaban dimensiones de ansiedad, evación y repetición de la experiencia respectivamente (*véase* tabla [17-9]. Sin embargo, en contraposición con las sub-categorías definidas en el DSM-IV, según las cuales la evación representa una dimensión de sintomatología, en esta muestra, la evación parecía estar separada en dos factores. El segundo factor reflejaba la evación relacionada con el repliegue general o el entorpecimiento de la sensibilidad, con altas cargas factoriales en los ítems "incapacidad de sentir emociones" y "menor interés en las actividades diarias". El cuarto factor reflejaba evación de estímulos relacionados con el o los hechos traumáticos (p. 104).

Explique los resultados a una persona que está familiarizada con la correlación pero no sabe nada acerca de análisis factorial.

4. Aron et al. (1998) realizaron un estudio acerca de las experiencias del amor no correspondido, es decir, amar a alguien que no nos ama. Una de las predicciones se concentró en la intensidad de la experiencia (cuánto piensa uno en ello, cuánto altera nuestras vidas). Los investigadores elaboraron la hipótesis de que la intensidad podría predecirse a través de la calidad de deseable de la relación (en qué medida la persona enamorada percibía que sería maravilloso tener una relación con la persona amada), la probabilidad (en qué medida el enamorado sentía que el ser amado lo había llevado a creer que podría desarrollarse una relación) y el deseo del estado (en qué medida el enamorado sentía que era deseable estar enamorado, aun cuando ese amor no fuera correspondido). Además, los investigadores plantearon la hipótesis de que el patrón de relación de las tres variables con la intensidad variaría conforme al estilo usual de vinculación afectiva

del enamorado (seguro, evasivo o ansioso-ambivalente, según lo tratado en el capítulo 11). Aron et al. realizaron un análisis de cada grupo a través del modelo de ecuación estructural. La figura 17-4 indica los resultados.

a) Explique el patrón de resultados. b) Utilizando este diagrama como ejemplo, explique los principios generales de la interpretación de un diagrama de senderos (incluso las limitaciones) a una persona que comprende la regresión múltiple en general pero no conoce los diagramas de senderos o los modelos de ecuación estructural.

5. Gire (1997) analizó los métodos preferidos para la resolución de conflictos, comparando personas de culturas individualistas con otras de culturas colectivistas. Los participantes eran 90 nigerianos (Nigeria fue considerada un ejemplo de sociedad relativamente colectivista) y 95 canadienses (Canadá fue considerada un ejemplo de sociedad relativamente individualista). Todos los participantes contestaron preguntas acerca de sus preferencias en cuanto a cada uno de cinco métodos de resolución de conflictos. La mitad de los participantes de cada país contestó las preguntas referidas a un conflicto interpersonal (un conflicto entre dos vecinos) y, la otra mitad, las relacionadas con un conflicto intergrupal (entre dos grupos de vecinos). El resultado del procedimiento fue un diseño factorial 2 (culturas) x 2 (conflicto interpersonal vs conflicto intergrupal), con cinco medidas de preferencias para la resolución de conflictos.

Los datos fueron analizados utilizando MANOVA. El MANOVA reveló en dos sentidos un efecto esencial significativo de la cultura $F(5, 173) = 6{,}37$, $p < 0{,}001$. El estudio del análisis univariado y de las medias sugiere que los nigerianos preferían la negociación mucho más que los canadienses, mientras que ocurría lo contrario con el arbitraje, conforme a lo que se había predicho. También hubo un resultado significativo de la cultura por tipo de interacción conflictiva, $F(5, 173) = 3{,}84$, $p < 0{,}002$. El análisis univariado y las medias, que se indican en la tabla [17-10], revelan que existieron diferencias significativas en tres procedimientos: amenazas, aceptación de la situación y arbitraje (p. 41).

Explique los resultados a alguien que comprende el análisis factorial de varianza pero no el análisis multivariado de varianza.

6. ¿Cuál sería la técnica estadística más apropiada para cada uno de los siguientes estudios ficticios?

a) Un estudio en el que el investigador sostiene una compleja teoría sobre el patrón de causa y efecto entre diversas variables.

b) Un estudio del grado de asociación entre dos variables.

c) Un estudio para determinar si una medida es internamente coherente y consistente a lo largo del tiempo en cuanto a dar el mismo resultado.

d) Un diseño factorial de 3 x 2 con tres variables dependientes.

e) Un estudio en el que se han medido siete variables que se consideran variables de predicción de determinada variable dependiente y el investigador desea determinar qué variables contribuyen significativamente a la predicción (pero no tiene ninguna teoría acerca de cuáles tienen mayores probabilidades de ser las más significativas).

f) Un estudio en el que el investigador mide 16 variables en una gran cantidad de participantes y desea averiguar si existen agrupaciones de variables implícitas más simple.

g) Un estudio en el que se comparan un grupo experimental y un grupo de control según una sola variable dependiente.

h) Un estudio que compara cinco grupos de individuos conforme a una sola variable dependiente.

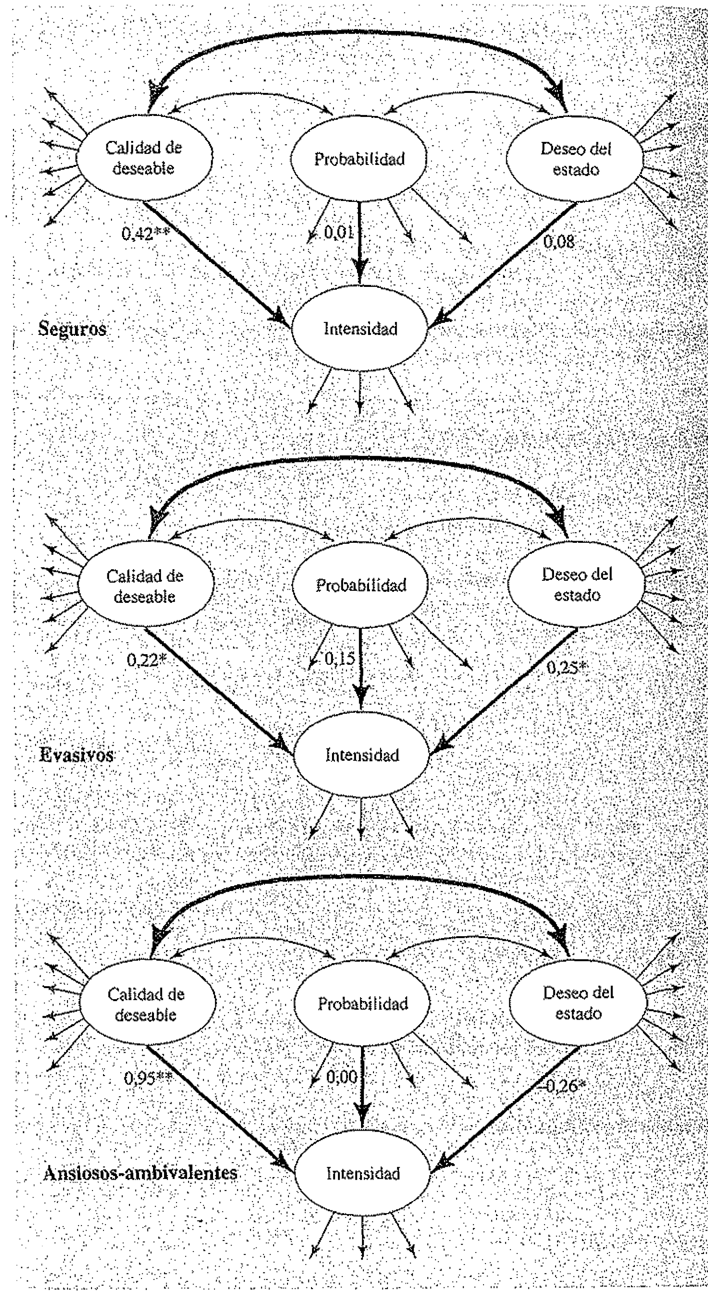
i) Un estudio en el que el investigador está analizando el efecto de diversas variables de predicción en una sola variable dependiente, tiene una teoría específica acerca de la importancia relativa de dichas variables, y desea verificar si cada variable de predicción agregada sucesivamente aporta algún elemento a la predicción lograda a través de las variables anteriores.

Tabla 17-9.
Cargas factoriales del análisis de los componentes principales de los síntomas de PTSD, según el DSM-IV, en 74 refugiados vietnamitas.

Carga **Dimensión sintomática**	**Factorial**
Ansiedad	
Pesadillas recurrentes	0,79
Dificultad para concentrarse	0,78
Irritabilidad/ataques de ira	0,77
Incapacidad a recordar partes de los hechos más traumáticos	0,74
Problemas de insomio	0,73
Evitar actividades que recuerdan el hecho traumático	0,70
Nerviosismo, facilidad para sobresaltarse	0,67
% de varianza explicada	44%
Evasión/abstinencia	
Incapacidad a sentir emociones	0,79
Menor interés en las actividades diarias	0,70
Sentimiento de indiferencia o abstinencia	0,65
Nerviosismo, facilidad para sobresaltarse	0,51
Sensación de que uno no tiene futuro	0,51
% de varianza explicada	24%
Repetición de la experiencia	
Pensamientos o recuerdos recurrentes de los hechos más terribles	0,83
Sensación de que el hecho está sucediendo nuevamente	0,83
Reacción emocional o física repentina cuando se le recuerdan los hechos más traumáticos	0,57
% de varianza explicada	22%
Evitar estímulos relacionados con el (los) hechos traumáticos(s)	
Evitar pensamientos o sentimientos relacionados con los hechos traumáticos	0,71
% de varianza justificada	11%

Fuente: Fawzi, M. C. S., et al. (1997), tab. 1. "Validez del estrés postraumático entre refugiados vietnamitas". *Revista Científica de Estrés Traumático [Journal of Traumatic Stress],* 10, 105. Copyright, 1997, por la Sociedad Internacional de Estudios del Estrés Traumático. Reimpreso con autorización.

Figura 17-4.
[Figura 2 de Aron et al. (en impresión). "Motivaciones para el amor no correspondido" *Boletín de Psicología social y de la personalidad. [Personality and Social Psychology Bulletin.]*



SERIE II

1. Aron & Aron (1997) realizaron un estudio concentrándose en las personas altamente sensibles a la estimulación. Los individuos mencionados tienden a descubrir sutilezas y notar cosas que otros pasan por alto, por lo cual puede encontrárselos en mayor medida entre los artistas y otros tipos de personas talentosas. Por otro lado, esa misma sensibilidad hace que estos individuos sufran, con más facilidad, de exceso de ansiedad. Lo que para las personas en general es un nivel normal de estimulación, con frecuendia resulta estresante para los individuos altamente sensibles. Aparentemente, como resultado de lo anterior, algunas PAS presentan niveles de emocionalidad (angustia y depresión) más altos que lo usual. Como parte del estudio en cuestión, los investigadores deseaban investigar si la sensibilidad era independiente de la emocionalidad. Por lo tanto, hicieron que un gran grupo de participantes completara cuestionarios acerca de sensibilidad y emocionalidad, junto con una serie de preguntas sobre diversas sensibilidades específicas y reacciones emocionales también específicas. Los investigadores estaban especialmente interesados en saber si la relación entre la escala PAS y diversas sensibilidades específicas permanecería aun después de controlar la emocionalidad en general; y además si la relación de la emocionalidad con reacciones emocionales específicas permanecería después de controlar la sensibilidad.

Como lo indica la tabla [17-11], la mayoría de las correlaciones entre las variables relacionadas con la sensibilidad y la escala PAS continuaron siendo significativas o casi-significativas después de excluir la medida de emocionalidad [...] Además [...] diversas variables pertinentes (p. ej. sentimientos emergentes) presentaban asociaciones únicas o exclusivas con la emocionalildad) (p. 354).

Explique el método y el resultado descriptos anteriormente a una persona que está familiarizada con la correlación y, en forma general, con la regresión múltiple común, pero que nunca ha oído hablar de la correlación parcial.

2. Shapiro et al. (1997) realizaron un estudio para desarrollar una medida de las actitudes de los niños hacia las armas y la violencia. La primera medida que desarrollaron tenía 61 ítems, e informaron que los análisis que reali-

Tabla 17-10.
Preferencias en cuanto a método como función de la cultura y el tipo de conflicto

	Nigerianos		Canadienses	
Método	*IP*	*IG*	*IP*	*IG*
Amenazas*	2,09	1,50	1,35	1,61
Aceptación dela situación*	2,72	3,16	3,43	2,71
Negociación	6,07	6,11	5,56	5,64
Mediación	4,70	4,77	4,87	5,13
Arbitraje*	3,05	4,90	5,20	5,42

Nota: Un asterisco (*) indica que las medias de la cultura por tipo de interacción conflictiva en cuanto a determinado método fueron significativas al nivel $p < 0,05$. A mayor número, mayor la preferencia por el método. IP (*Interpersonal Conflict*, Conflicto Interpersonal); IG (*Intergrupal Conflict*, Conflicto intergrupal).

Fuente: Gire, J. T. (1997), tab. 1. "El efecto variante del individualismo-colectivismo con respecto a los métodos preferidos para la resolución de conflictos". *Revista Científica Canadiense de la Ciencia del Comportamiento [Canadian Journal of Behavioural Science]*, 29, 38-43. Copyright, 1997, por la Asociación Canadiense de Psicología. Reimpreso con autorización.

zaron "indican un nivel altamente satisfactorio de coherencia interna del cuestionario... (el α de Cronbach = 0,94)" (p. 314). Con el fin de crear una medida más breve y práctica, redujeron la escala a 23 ítems, y luego explicaron: "Realizamos diversos analisis para determinar si la disminución de la longitud se obtenía a costa de perder coherencia interna... el alfa de Cronbach de la medida reducida fue de 0,88 (vs 94)" (p. 314). Explique los resultados descriptos a alguien familiarizado con la correlación pero no con la confiabilidad o el alfa de Cronbach.

Tabla 17-11.
Correlaciones y correlaciones parciales de la sensibilidad y la emocionalidad con variables relacionadas con la sensibilidad. Estudios 2-4.

			Correlaciones parciales	
Variable	*Escala PAS*	*Emocionalidad*	*Escala PAS (emocionalidad)*	*Emocionalidad (escala PAS)*
Estudio 2				
Llora con facilidad	0,36**	0,38**	0,21**	0,24**
Sensibilidad a la luz del día	0,32**	0,26**	0,25**	0,11*
Sensibilidad al alcohol	0,39**	0,18**	0,36**	–0,03
Prefiere la música *country* en vivo	0,22**	0,08	0,22**	–0,04
Las películas lo afectan al día siguiente	0,31**	0,23**	0,23**	0,10†
Intensidad en el amor	0,26**	0,30**	0,14*	0,19**
Sentimeintos emergentes[a]	0,28**	0,30**	0,18**	0,17**
Recuerda los sueños[b]	0,19*	0,03	0,20†	–v08
Sueños intensos[b]	0,19*	0,08	0,18†	–0,03
Tiempo en soledad[b]	0,22*	0,07	0,17†	–0,05
Estudio 3				
Llora con facilidad	0,47**	0,46**	0,27**	0,26**
Prefiere la música *country* en vivo	0,15*	0,10†	0,11†	–0,00
Las películas lo afectan al día siguiente	0,30**	0,17**	0,11†	0,08
Intensidad en el amor	0,23**	0,16**	0,17**	0,03
Estudio 4				
Prefiere la música *country* en vivo	0,09*	0,07	0,07	0,03
Sueños intensos	0,19**	0,05	0,18**	–0,03

Nota: el estudio 2 incluyó 313 alumnos de la Universidad de California, Santa Cruz; el estudio 3 incluyó datos de 285 alumnos universitarios norteamericanos no graduados; el estudio 4 incluyó datos tomados de 301 personas a través de una encuesta telefónica pública de discado aleatorio. PAS = Persona altamente sensible.

[a] El ítem así señalado ("¿Le surgen sentimientos muy intensos sin razón aparente?") fue contestado sólo por 211 participantes.

[b] Los ítems así señalados fueron completados sólo por 107 participantes.

$*p < 0,05$; $**p < 0,01$; $\dagger\, p < 0,10$.

Fuente: Aron, E. N., & Aron, A. (1997), tab. 3. "Sensibilidad del proceso sensorial y su relación con la introversión y la emocionalidad". *Revista Científica de Psicología Social y de la Personalidad [Journal of Personality and Social Psychology]*, 73, 345–368. Copyright, 1997, por la Asociación Americana de Psicología. Reimpreso con autorización.

3. Crick et al. (1997) realizaron un estudio para desarrollar una medida, por parte de maestros, de la "agresión a través de relaciones" en niños de edad preescolar. Comúnmente, la agresión manifiesta daña a otros directamente, pero la "agresión a través de relaciones" daña a otros a través del perjuicio a las relaciones de éstos con sus pares (p. ej. utilizando la exclusión social o esparciendo rumores como una forma de represalia) (p. 579).

Tabla 17-12
Cargas factoriales de la medida de comportamiento social evaluado por maestros (PSBS–T).

Ítem	Agresión a través de relaciones	Agresión manifiesta	Comportamiento prosocial	Alteración depresiva
Le informa a un compañero que no jugará con él ni será su amigo a menos que haga lo que él le pide	0,84			
Ordena a otros que no jueguen con algún compañero o que no sean sus amigos	0,83			
Cuando se enoja con un compañero, el niño hace que ese compañero no pueda estar con el grupo de amigos de juegos	0,81			
Amenaza a un compañero o compañera diciéndole que no será invitado/a a las fiestas de cumpleaños a menos que haga lo que él quiere	0,88			
Intenta que otros sientan antipatía por un compañero	0,89			
Amenza verbalmente con dejar a un compañero fuera del grupo de amigos de juegos si ese compañero no hace lo que el niño pide	0,85			
Patea o golpea a otros		0,81		
Amenaza verbalmente con pegar o golpear a otros niños		0,75		
Arruina las cosas de sus compañeros cuando está enojado o enojada		0,82		
Suele empujar a otros niños		0,72		
Daña a otros niños pellizcándolos		0,83		
Amenaza verbalmente con causar daño físico a un compañero para obtener lo que desea		0,81		
Sabe compartir y turnarse			0,76	
Es servicial con sus compañeros			0,83	
Es amable con sus compañeros			0,62	
Dice o hace cosas lindas a otros niños			0,75	
No se divierte mucho				0,90
Parece triste				0,87
Sonríe poco				0,82

Nota: Todas las cargas cruzadas eran menores a 0,40. PSBS–T (*Preschoal Behavoir Scale -Teacherform*, Escala de comportamiento social preescolar, formulario para el maestro).

Fuente: Crick, N. R., Casas, J. F., & Mosher, M. (1997), tab. 1. "Agresión manifiesta a través de relaciones en el preescolar". *Psicología para el Desarrollo, Developmental Psychology*, 33, 579-588. Copyright, 1997, por la Asociación Americana de Psicología. Reimpreso con autorización.

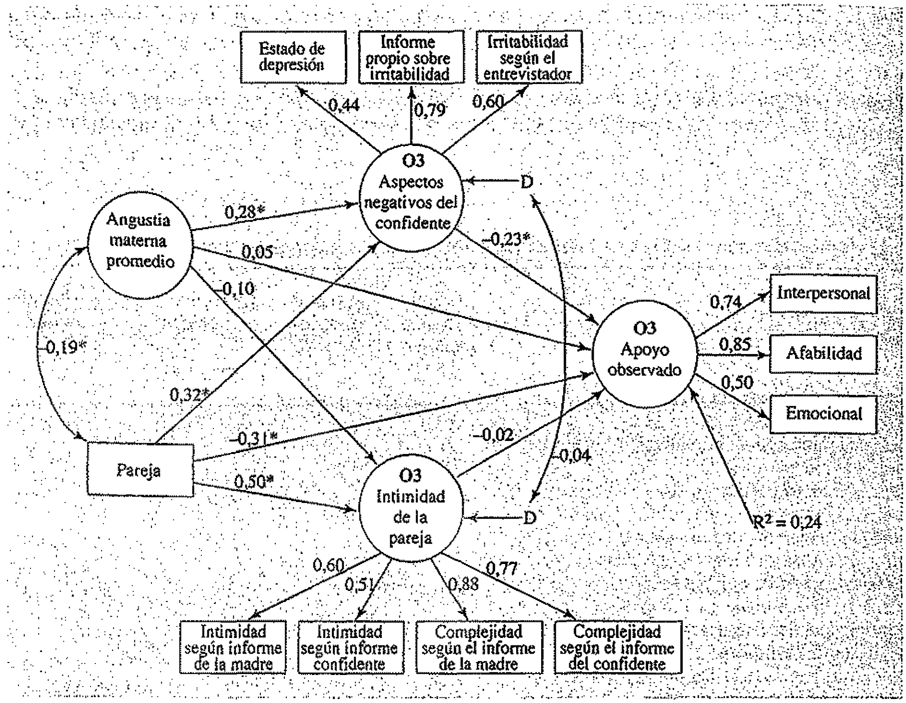


Figura 17-5.
Diseño de proceso de apoyo con características de la madre, del confidente y de la relación, controlando el hecho de la formación de una nueva pareja con un hombre y el cambio en la angustia materna. O3 = 3° oportunidad; Conf. = Confidente. $x^2(67, N = 138) = 84,82$, $p = 0,07$; índice de concordancia comparativo = 0,963: $*p < 0,05$. [**Fuente**: DeGarmo, D. S., & Forgatch, M. S. (1997), fig. 1. "Determinantes del respaldo observado en el confidente hacia las madres divorciadas". *Revista Científica de Psicología Social y de la Personalidad [Journal of Personality and Social Psychology]*, 72, 336-345. Copyright, 1997, por la Asociación Americana de Psicología. Reimpreso con autorización]

Como parte del estudio, en primer lugar formularon una escala con 23 ítems para la valoración por parte de los maestros del comportamiento social de niños en edad preescolar. Los investigadores describieron de la siguiente manera el análisis principal de la medida mencionada:

> En primer lugar, se realizó un análisis factorial de componentes principales [...] para evaluar si [...] la agresión a través de relaciones surgiría como un factor separado independiente de la agresión manifiesta. El análisis produjo los cuatro factores predichos: agresión a través de relaciones, agresión manifiesta, comportamiento prosocial y alteración depresiva (p. 582).

La tabla 17-12 indica las cargas factoriales. Explique los resultados a una persona que está familiarizada con la correlación pero no conoce el análisis factorial.

4. DeGarmo y Forgatch (1997) realizaron un estudio acerca del apoyo social recibido por madres divorciadas de parte de sus confidentes más cercanos. Como parte del estudio, midieron una cantidad de variables y después analizaron las relaciones predichas entre las variables, utilizando el modelo de ecuación estructural. La figura 17-5 representa gráficamente los resultados.

a) Explique el patrón de resultados. b) Utilizando como ejemplo el diagrama presentado, explique los principios generales de la interpretación de un diagrama de senderos (incluso las limitaciones) a una persona que comprende la regresión múltiple en líneas generales pero no conoce los diagramas de senderos o el modelo de ecuación estructural.

5. En la biblioteca, busque en una publicación reciente de alguna revista científica especializada en un área de la psicología, algún artículo que le interese especialmente y en el que se aplique uno de los procedimientos estadísticos descriptos en este capítulo. Redacte un breve resumen del estudio que encontró refiriéndose específicamente a los cálculos estadísticos. Con su respuesta incluya una fotocopia de la publicación, marcando claramente las partes en las que se informan los procedimientos estadísticos por usted descriptos.

6. En la biblioteca, busque en una publicación reciente de alguna revista científica especializada en un área de la psicología, algún artículo que le interese especialmente y en el que se aplique un procedimiento estadístico que no haya sido tratado en este libro. Redacte un breve resumen del estudio que encontró refiriéndose específicamente a los cálculos estadísticos. Con su respuesta incluya una fotocopia de la publicación, marcando claramente las partes en las que se informan los procedimientos estadísticos por usted descriptos.

APÉNDICE

A Repaso de la lógica y de la terminología relacionadas con la investigación psicológica

Los métodos estadísticos son herramientas utilizadas en el proceso de investigación. Los procedimientos estadísticos tratados en este libro resultarán más fáciles de comprender si se tiene en cuenta el contexto más amplio en el cual se insertan.

En la mayoría de los casos, el propósito de un estudio de investigación psicológica consiste en evaluar la validez de una teoría o la efectividad de una aplicación práctica[1]. El investigador puede adoptar muchos métodos. Los procedimientos de investigación más sólidos llevan a conclusiones inequívocas referidas a una amplia gama de situaciones y personas. Los diseños de investigación deficientes, aun cuando sus resultados sean coherentes con las predicciones del investigador, dejan abiertas muchas interpretaciones alternativas con respecto a los motivos por los cuales se llegó a ese resultado, o bien se aplican sólo a un reducido grupo de personas y situaciones.

A veces, las circunstancias limitan el tipo de procedimiento de investigación aplicable y, aún así, vale aparentemente la pena continuar con la investigación, incluso de un modo menos riguroso. De hecho, especialmente en el caso de las investigaciones aplicadas, muchos de los trabajos más importantes han sido realizados por psicólogos que utilizan (por necesidad) métodos menos que perfectos, pero muy creativos.

Sin embargo, la mayoría de los psicólogos analizan la lógica de la investigación en función de un tipo de método ideal. Por lo tanto, un estudio real se evalúa según las diferentes formas en las que se aproxima o deja de aproximarse a ese ideal. En este apéndice trataremos primero ese ideal

[1] Algunas veces se realizan investigaciones con otros fines, tales como explorar relaciones entre varias medidas, determinar la incidencia de alguna característica de la población, o desarrollar una medida o técnica para utilizar en otra investigación. Sin embargo, la lógica básica de la forma usual de investigación (tema central de este apéndice) sirve de apuntalamiento del modo en que los psicólogos abordan la mayoría de las investigaciones sistemáticas.

(el "verdadero experimento"), la terminología clave relacionada con él y, por último, nos dedicaremos a cuatro áreas clave en las que los estudios se aproximan o no a ese ideal: equivalencia de participantes entre grupos experimentales, equivalencia de circunstancias entre grupos experimentales, legitimidad de la generalización y suficiencia de la medición.

EL MÉTODO DE INVESTIGACIÓN TRADICIONALMENTE IDEAL

El experimento verdadero

El procedimiento de investigación que usualmente conduce al menor nivel de ambigüedad es el **experimento verdadero**. Es el estándar con el que se comparan todos los otros métodos. Partiendo de la hipótesis "cambiar el nivel de X provoca un cambio en el valor de Y", el experimento real varía sistemáticamente el nivel de X, manteniendo igual todos los demás aspectos, y observando el efecto en Y. Por ejemplo, supongamos que un investigador está interesado en averiguar si el hecho de que haya luces centelleantes en el aula afecta las calificaciones de las personas en una prueba de matemática, en donde X representa la existencia de luces centelleantes en el aula e Y las calificaciones en la prueba de matemática. En un experimento real, se tomaría la prueba a cada alumno de un determinado grupo en un aula con luces centelleantes. A otro grupo de alumnos, inicialmente idéntico, se le tomaría la prueba bajo condiciones completamente idénticas, pero sin la presencia de luces centelleantes en el aula. Así, la única diferencia entre los dos grupos sería el nivel de X, es decir, la presencia o ausencia de luces centelleantes en el aula. Si los alumnos del aula con luces centelleantes obtienen calificaciones menores en la prueba de matemática (Y), la causa tiene que ser la iluminación. (Si obtienen mejores calificaciones, también sería a causa de la iluminación).

Terminología básica del experimento

Gran parte de la terminología de investigación proviene del método que describimos anteriormente. Un grupo al que se manipula el nivel de X se lo denomina usualmente **grupo experimental.** El grupo de comparación en el que X se mantiene en niveles normales se lo denomina **grupo control**. Los individuos analizados en la investigación se llaman **participantes**[2]. La variable que se modifica sistemáticamente (X, por ejemplo, si las luces centellean o no) se denomina **variable independiente**. El procedimiento de modificación sistemática de la variable independiente a veces recibe el nombre de **manipulación experimental** o **manipulación de la variable independiente**. La variable que se supone que cambia como resultado del estudio (Y, si X es la causa de Y, por ejemplo la calificación en la prueba de matemática) se la llama **variable dependiente**. Los participantes a seleccionar, es decir, la **población**, constituyen el grupo que incluye a todas las personas que pertenecen al tipo bajo análisis. Aquellos miembros seleccionados entre la población conforman la **muestra** a analizar.

[2] Con frecuencia, los psicólogos utilizan el término "sujeto". Sin embargo, nosotros utilizamos la palabra "participante", aquí y a lo largo de todo el libro.

Como ejemplo, imaginemos que un investigador tiene dos latas idénticas de gaseosa. La hipótesis que se plantea para este caso es: "Al calentar una lata de gaseosa, ésta explotará".

(¡No se debe probar el experimento en casa!). En otras palabras, el aumento de calor causará una explosión. El investigador podría poner un fósforo bajo una lata (la lata experimental) y no ponerlo bajo la otra (la lata control). Si la lata experimental explota y la lata de control no, se confirma la hipótesis. Cada lata es un participante; el calentamiento es la variable independiente; la explosión de la lata es la variable dependiente, y las dos latas son las muestras, respectivamente, de las poblaciones de todas las latas de gaseosas calentadas y no calentadas (*véase* figura A-1).

Cuatro características del diseño de investigación ideal

Un diseño de investigación ideal presenta cuatro características clave:

1. Los participantes de los grupos experimental y de control son idénticos.
2. Los grupos experimentales y de control son expuestos a situaciones idénticas (**excepto** por la manipulación de la variable independiente).
3. La muestra analizada representa perfectamente la población objeto del estudio.
4. La medición de la variable dependiente es completamente precisa y adecuada para lo que se supone que está midiendo.

El resto del presente apéndice analiza las diversas formas en las que la investigación real intenta aproximarse a cada una de las condiciones ideales mencionadas.

EQUIVALENCIA DE PARTICIPANTES EN LOS GRUPOS CONTROL Y EXPERIMENTAL

Comúnmente, lo primero que se tiene en cuenta al evaluar si los resultados de un estudio llevan a conclusiones inequívocas es la equivalencia de participantes en los grupos control y experimental. Por ejemplo, supongamos que no estuviéramos seguros de que la capacidad en matemática de los miembros del grupo en el aula con las luces centelleantes fuera inicialmente la misma que la de aquellos en el aula sin luces centelleantes. Por lo tanto, cualquier diferencia en las calificacio-

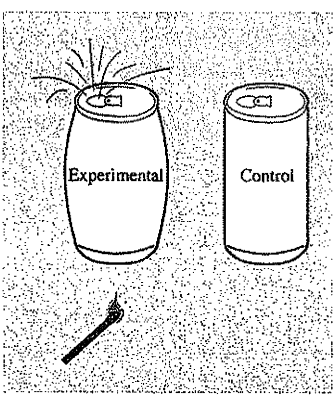


Figura A-1.
Un experimento ideal: se calienta una de dos latas de gaseosa idénticas, y el investigador observa si explota mientras que la otra no lo hace.

nes matemáticas entre los dos grupos, al finalizar el estudio, tendría un significado ambiguo. La diferencia podría ser el resultado de a) la manipulación de la variable independiente (tener o no luces centelleante), o bien de b) las diferencias iniciales en cuanto a la capacidad. Para evitar tales resultados ambiguos, los investigadores buscan una equivalencia estricta entre los grupos control y experimental. Se emplean cinco estrategias principales: asignación aleatoria a los grupos, diseño de grupo control equivalente, diseño de medidas repetidas, diseño de investigación correlacional, e investigación de sujeto único.

Asignación aleatoria a los grupos

El procedimiento científico que produce la aproximación real más cercana a dos grupos idénticos se denomina **asignación aleatoria a los grupos**. Por ejemplo, si había 100 personas disponibles para participar en un experimento, cada persona podría incluirse tanto en el grupo experimental como en el control lanzando una moneda. Aunque los dos grupos de 50 personas creados del modo descripto no son idénticos, al menos no habrá diferencia **sistemática** entre ellos.

Es importante recalcar que "aleatoria" significa utilizar un procedimiento de selección estrictamente de azar, y no simplemente elegir personas sin ningún plan u orden establecido para incluirlas en los dos grupos, ya que todo procedimiento sin planificación ni orden alguno puede producir diferencias sistemáticas no intencionales. Por ejemplo, si el investigador hubiera elegido un grupo entre los alumnos que asisten a una clase matutina y el otro entre los que asisten a una clase vespertina, los dos grupos podrían diferir entre sí, ya que los tipos de personas que asisten a clases en esos distintos horarios podrían ser diferentes. O bien, supongamos que un grupo está formado por voluntarios que se ofrecen para realizar un programa de mejora de la autoestima, y el grupo de control está formado simplemente por quienquiera que esté dispuesto a realizar una *prueba de autoestima*. Los tipos de personas que forman el grupo experimental y el de control podrían ser bastante diferentes.

La asignación aleatoria excluye las diferencias sistemáticas iniciales entre los grupos. Cualquier diferencia efectiva que exista después de la asignación aleatoria será causa enteramente de los procesos aleatorios. En consecuencia, si después del experimento se revelan diferencias en la variable dependiente, dichas diferencias sólo pueden ser el resultado de la manipulación de la variable independiente o bien de los procesos aleatorios. Los verdaderos procesos aleatorios cumplen las leyes de la probabilidad, por eso los procedimientos de prueba de hipótesis tratados a partir del capítulo 6 de este libro pueden verificar la probabilidad de que la diferencia encontrada en un estudio pudiera haber sido el resultado de los procesos aleatorios. Si el análisis estadístico indica que esa posibilidad es improbable, la única explicación razonable restante es que la diferencia haya sido el resultado de la manipulación de la variable independiente. Lo que acabamos de describir es la lógica básica que está implícita en el análisis de los resultados experimentales, la cual explica por qué la asignación aleatoria y los métodos estadísticos son tan importantes en la investigación psicológica.

Diseños de grupo de control equivalente

A veces, la asignación aleatoria a los grupos no es práctica. Por ejemplo, la ética exigiría que todos los alumnos de un distrito escolar, que necesiten determinado programa de lectura, lo reciban; por lo tanto, no puede decidirse al azar que algunos pierdan esa oportunidad. ¿Cómo podemos demostrar entonces que el programa es la causa de mejoras en el alumno? Un método de investigación alternativo ampliamente utilizado es el **diseño de investigación con grupo de control**

equivalente. Por ejemplo, un investigador podría comparar un grupo experimental de alumnos que han sido seleccionados para el programa en un distrito escolar, con un grupo de control formado por alumnos de otro distrito, que también necesitan el programa, pero para quienes el mismo no se encuentra disponible. Cada miembro de ese grupo de control podría ser equiparado a un miembro del grupo experimental en cuanto a edad, clase social, sexo, problemas de lectura, y así sucesivamente.

Los diseños de grupo de control equivalente resultan ser mucho mejores que carecer de grupo de control (de hecho, si se prueban ambos grupos antes y después, el diseño de grupo de control equivalente puede dar resultados relativamente inequívocos). La situación descripta, denominada **diseño con grupo de control equivalente y pruebas previa y posterior**, es un ejemplo de **diseño cuasiexperimental**. Un diseño cuasiexperimental es aquel método que se aproxima razonablemente a un experimento verdadero, pero que no utiliza la asignación aleatoria.

Sin embargo, no importa cuán adecuada sea la equiparación entre un par de grupos, ya que incluso utilizando pruebas previas y posteriores, un investigador nunca puede estar seguro de que no existe diferencia inicial sistemática entre los grupos. En efecto, en la mayoría de los casos, si no se ha utilizado asignación aleatoria, se sabe que **existe** una diferencia inicial sistemática, cualquiera sea el elemento que ubicó a las personas en uno u otro grupo. (En el ejemplo relacionado con el programa de lectura, la diferencia sistemática podría ser que un grupo de alumnos viva en un distrito escolar que no es progresista o que no cuente con los fondos suficientes para ofrecer el programa de lectura).

Diseños de medidas repetidas

Otro método de investigación es crear dos grupos idénticos probando a las mismas personas dos veces. Este procedimiento se conoce como **diseño de investigación de medidas repetidas** (también se lo denomina **diseño de investigación intrasujeto**). Los alumnos del ejemplo mencionado podrían ser puestos a prueba antes y después del programa de lectura.

El diseño de medidas repetidas más simple es el **diseño de grupo único con prueba previa y posterior**, en el cual, como su nombre lo indica, se prueba dos veces un sólo grupo de individuos, una vez antes y otra vez después de algún tratamiento experimental. No obstante, este tipo de diseño de investigación resulta muy débil en cuanto a que, si se descubre un cambio, hay muchas explicaciones alternativas posibles para el mismo. Simplemente el hecho de ser probado por primera vez puede cambiar a un participante, de modo tal que cuando vuelve a ser puesto a prueba, esa persona no es idéntica sino diferente (es diferente a causa de la prueba inicial, no por tratamiento experimental). Incluso el tiempo mismo produce cambios. En líneas más generales, en el tipo de estudios mencionados cualquier cambio podría ser el resultado del programa de lectura o de cualquier otra cosa que hubiera sucedido a los participantes durante ese período (además del tratamiento experimental). O bien podría haber tendencias preexistentes hacia la mejora, o bien el cambio podría ser el resultado de un proceso general de maduración y experiencia, o bien podría ser que la persona hubiera comenzado en un nivel muy bajo de modo que mejoraría naturalmente sin el tratamiento, y así sucesivamente.

Debido a que se trata de un diseño de investigación tan débil, el diseño de grupo único con prueba previa y posterior se considera **diseño preexperimental**. Se trata de un tipo de investigación que en general tiene suma importancia como primer paso en la exploración de un campo de investigación, pero cualquier conclusión derivada de un estudio de este tipo sería muy tentativa, para lo cual debería ser seguida por un diseño de investigación más sólido (como por ejemplo, un diseño cuasiexperimental o un experimento real).

Sin embargo, dentro del entorno del laboratorio, se utiliza con frecuencia un diseño de medidas repetidas, de forma tal que se lo transforma en un experimento real. Supongamos que un investigador está interesado en el efecto que causa la iluminación en el rendimiento de una tarea complicada. El investigador podría probar el rendimiento de las distintas personas bajo luces brillantes (condición experimental), y luego probar nuevamente el rendimiento de esas personas bajo iluminación normal (condición de control). Sin embargo, un inconveniente que presenta este método es que los participantes podrían estar más familiarizados con la prueba en la segunda oportunidad en que la realicen, **creándose así un efecto de práctica o traspaso**; o bien, podrían estar cansados para el momento en el que les tocara hacer la segunda prueba, creándose un **efecto de fatiga**. Para resolver ese tipo de problemas, los investigadores utilizan un procedimiento denominado **compensación**, en el cual la mitad de los participantes se prueban primero según una condición, y la otra mitad se prueba primero según la otra condición. De ese modo, cualquier efecto de práctica, traspaso, fatiga o similar se compensa entre las dos condiciones. Lo ideal sería que la compensación se emplee de modo que se asigne en forma aleatoria la condición que un participante experimentará primero y, en ese caso, el estudio se transforma en un experimento real. De hecho, debido a que logra un nivel de equivalencia tan importante entre los grupos, el diseño de medidas repetidas con compensación y asignación aleatoria es uno de los métodos de investigación más poderosos que utilizan los psicólogos.

Diseño de investigación correlacional

Un **diseño de investigación correlacional** examina el grado de asociación entre dos variables tal como existen en un grupo de personas, sin ningún intento de manipulación experimental. Así, un método correlacional para analizar la autoestima y la satisfacción laboral consistiría simplemente en sondear a un grupo de gerentes de nivel medio en cuanto a su autoestima y a su satisfacción laboral. Luego, el investigador observaría si aquellos con valores altos en cuanto a autoestima presentaron, por lo general, valores altos en cuanto a la satisfacción laboral. (El grado en el que efectivamente existe una asociación entre los dos valores se calcula utilizando una técnica estadística denominada "coeficiente de correlación", descripta en el capítulo 3).

El método correlacional es, con frecuencia, el más adecuado para las circunstancias, por lo cual es ampliamente utilizado. Pero resulta un diseño de investigación bastante débil en cuanto a que sus resultados están sujetos a muchas explicaciones alternativas, además de que "*X* fue la causa de *Y*". Por ejemplo, supongamos que a través de un estudio correlacional se descubre que la autoestima y la satisfacción laboral están relacionadas. El resultado podría reflejar que la alta autoestima causa alta satisfacción laboral. Sin embargo, también podría ser que un alto grado de satisfacción laboral cause un alto grado de autoestima. La relación entre autoestima y satisfacción laboral podría incluso ser el resultado de otras diferencias entre los gerentes, tales como la edad (tal vez el ser mayores hace que los gerentes de nivel medio tengan un alto nivel de autoestima como de satisfacción laboral). (En el capítulo 3, tratamos con cierto grado de detalle las diversas interpretaciones causales de los resultados de un estudio correlacional). Así, una ventaja del experimento verdadero (cuando es factible), con respecto al estudio correlacional, es que el experimento verdadero manipula la variable independiente y luego observa el efecto en la variable dependiente, indicando con bastante claridad cuál es la causa y cuál el efecto.

Los investigadores son muy conscientes de los límites de los diseños correlacionales. Cuando es posible, intentan anular algunas explicaciones alternativas utilizando principalmente procedimientos estadísticos sofisticados, tales como la correlación parcial (resumida en el capítulo 17). Aun así, el método correlacional nunca produce resultados tan inequívocos como los de un experimento verdadero y, en la mayoría de los casos, ni siquiera tan bien definidos como los de un cuasiexperimento. No obstante, es el método de investigación más sólido que puede resultar viable en una gran cantidad de situaciones. Por ejemplo, uno no puede realizar una asignación aleatoria para que ciertas personas contraigan matrimonio con otro tipo de personas. Incluso cuando es posible realizar experimentos pueden resultar muy costosos y, en esos términos, los investigadores pueden no querer o no poder investigar experimentalmente una idea no probada. En casos como los mencionados, los estudios correlacionales constituyen, con frecuencia, un importante primer paso para dar a conocer una nueva área de investigación.

Investigación de sujeto único

Por último, algunos estudios de investigación involucran en un intenso examen a un sólo grupo, organización o individuo, utilizando el método de "análisis de un caso" u "observación participante". Tal **investigación de sujeto único** no se considera experimental, ni siquiera correlacional. No obstante, en el área de la psicología clínica y en algunos otros campos de las ciencias sociales como la sociología y la antropología (y dentro de éstos, los enfoques sociológicos y antropológicos del comportamiento corporativo, de la educación, de la criminología, de la comunicación, etc.), este tipo de investigaciones son consideradas valiosas ya que permiten comprender acabadamente toda la complejidad de lo que se analiza, en lugar de forzar la atención hacia unas pocas variables que pueden o no ser las más críticas. En todos los campos de la psicología, así como también de las otras ciencias sociales, la investigación de sujeto único se considera valiosa como precursora de otros métodos de investigación más rigurosos. (El capítulo 2 trata los temas mencionados).

Los investigadores también utilizan el método de sujeto único en forma sumamente sistemática en la tradición conductista desarrollada por B. F. Skinner. Se analiza un sólo sujeto a lo largo del tiempo, ya sea un animal –como una rata o una paloma– o un paciente que sigue un programa de terapia conductista, mientras el investigador sistemáticamente manipula las condiciones que afectan al participante y observa los cambios resultantes. Generalmente no utilizan la estadística, el patrón de resultados debe ser lo suficientemente claro como para que la estadística sea innecesaria.

Resumen de los distintos diseños de investigación

La tabla A-1 resume los distintos diseños de investigación que hemos examinado, señalando sus ventajas y desventajas en comparación con el ideal de grupos experimental y de control idénticos.

EQUIVALENCIA DE CIRCUNSTANCIAS EN LOS GRUPOS EXPERIMENTAL Y DE CONTROL

El estudio ideal no sólo requiere grupos idénticos sino también que las circunstancias de prueba sean idénticas.

En la práctica, es bastante difícil probar dos grupos bajo circunstancias en las que la única diferencia es la manipulación de la variable independiente. En un laboratorio de física es posible lograr esa equivalencia, pero al realizar investigaciones con humanos las circunstancias nunca son equivalentes. Una estrategia diseñada para maximizar la equivalencia es utilizar un lugar aislado, como por ejemplo un compartimiento de un edificio de estudios psicológicos, minimizando las influencias externas e interrupciones que podrían hacer que una sesión del experimento fuera di-

Tabla A-1.
Principales diseños de investigación, sus ventajas y desventajas.

Diseño	Ventajas	Desventajas
Experimento verdadero (asignación aleatoria a las condiciones).	Asegura la ausencia de diferencias sistemáticas entre las condiciones.	Su implementación puede no ser viable o contraria a la ética.
Grupo control equivalente (sin asignación aleatoria).	Controla diferencias obvias entre condiciones. Puede ser el más práctico con grupos intactos.	Los grupos pueden diferir sistemáticamente con respecto a variables en las que no han sido equiparados.
Grupo control equivalente con prueba previa y posterior.	Controla con bastante fuerza las diferencias iniciales entre participantes. Con frecuencia resulta práctico cuando la asignación aleatoria no lo es.	Las diferencias sistemáticas entre los grupos pueden influir en el impacto. El procedimiento de medición previo a la prueba puede confundir los resultados.
Experimento verdadero de medidas repetidas (asignación aleatoria).	Asegura la ausencia de diferencias sistemáticas. Minimiza las diferencias aleatorias convirtiendo a los participantes en sus propios controles.	Efectos de práctica o traspaso. El procedimiento puede ser difícil de implementar.
Grupo único con prueba previa y posterior.	Brinda cierto control. Con frecuencia es el único método viable.	Es imposible saber si el cambio hubiera ocurrido sin el tratamiento experimental.
Correlacional.	Es relativamente fácil de implementar con grupos intactos.	Dificultad para determinar la dirección de causalidad.
Sujeto único.	Permite un conocimiento acabado de los procesos.	Dificultad para generalizar los resultados.

ferente de otra. Un método relacionado con el tema mencionado consiste en estandarizar la situación al máximo; por ejemplo, las instrucciones para los participantes podrían estar grabadas.

Sin embargo, con respecto a la equivalencia de circunstancias existen dos inconvenientes especiales que condicionan la mayoría de las investigaciones de las ciencias sociales, particularmente la investigación aplicada: nos referimos a los efectos del experimentador y a los efectos placebo o Hawthorne.

Efectos del experimentador

Los **efectos del experimentador**, que incluyen el **sesgo del experimentador**, son las influencias no intencionales del investigador sobre el estudio. Por ejemplo, supongamos que en un estudio acerca de los efectos de la terapia psicológica, el investigador es un terapeuta que evalúa la salud mental de los participantes. En ese caso, es muy probable que el deseo del terapeuta de que el experimento funcione lo predisponga a observar que los participantes en el grupo experimental han mejorado en mayor medida. Incluso, si un observador independiente clasificara los dos grupos sabiendo quién forma parte de cada uno, el deseo de que el experimento resulte de determinada manera podría influir no intencionalmente en las evaluaciones del observador.

La mejor solución para el inconveniente descripto se denomina **método de prueba a ciegas**; es decir, el experimentador, al momento de interactuar con el participante, no sabe si éste se encuentra en el grupo de control o en el experimental.

Efectos placebo y Hawthorne

Los **efectos placebo** influyen sobre la expectativa o motivación de un participante con respecto a desempeñarse adecuadamente. Los **efectos Hawthorne** influyen sobre la atención que recibe el participante y sobre la reacción del mismo por su condición de tal. Por ejemplo, si en una fábrica se capacita al personal de un sector en un nuevo programa y al personal de otro sector no, los dos grupos presentan varias diferencias. Un sector utiliza la nueva forma de operación resultante del programa y el otro sector no; esa es la manipulación de la variable independiente. Sin embargo, también existe otra diferencia, y es que a un sector se lo ha inducido a esperar beneficios y al otro sector no (creándose un efecto placebo). Otra diferencia es que un sector ha recibido atención especial y el otro sector no (creándose un efecto Hawthorne: el término proviene de un estudio realizado en el año 1927 en la Planta Industrial de Hawthorne de la Western Electric Company, en Cicero, Illinois). Las diferencias adicionales entre los grupos confunden en gran medida la interpretación del efecto causado por la manipulación de la variable independiente.

¿Cómo pueden remediar los investigadores estas diferencias circunstanciales no deseadas? La mejor solución es realizar un estudio en el que ambos grupos reciban cierto tratamiento, y que además estén convencidos de que debería ser provechoso. Sin embargo, sólo un grupo recibe efectivamente un tratamiento que incluye algo más que una mera atención y mayores expectativas. Por ejemplo, en investigaciones médicas, ambos grupos recibirían píldoras iguales y con el mismo sabor, pero las píldoras de un grupo contienen el ingrediente activo, y las del otro grupo no. Ninguno de los participantes del experimento sabe quién recibe la droga real. Entonces una droga que se ve y que se sabe que es la droga verdadera, pero que en verdad es inactiva, se denomina un **placebo** (en latín "complacer").

En psicología, generalmente es imposible o contrario a la ética establecer un grupo de control en el que una persona reciba un tratamiento que cree efectivo pero en realidad no lo es. Aquellas situaciones en las que es factible utilizar un verdadero grupo de control placebo y en el que incluso el personal relacionado con la investigación desconoce qué participantes pertenecen a cada grupo, se denominan **procedimientos doble ciego**.

Los efectos placebo y Howthorne son los inconvenientes más comunes que se presentan al intentar sacar conclusiones inequívocas a partir de los resultados de investigaciones aplicadas en áreas tales como la psicología clínica, educativa y empresarial.

REPRESENTATIVIDAD DE LA MUESTRA

El tercer requisito para lograr un estudio ideal es que la muestra de participantes analizados represente *adecuadamente la población a la que se supone que se aplica el estudio*. Esa *representatividad* se denomina **legitimidad de la generalización** o **validez externa**. (La **validez interna** se refiere a las cuestiones relacionadas con la equivalencia de los grupos experimental y de control y a la equivalencia de circunstancias).

La investigación psicológica se realiza con frecuencia en alumnos universitarios, y se supone que lo que se descubre acerca de ellos se aplica a la población más amplia formada por las personas en general. En un estudio sobre el efecto que producen las luces centelleantes en el desempeño, el patrón general de resultados con alumnos universitarios probablemente se aplique a casi todos los otros seres humanos. No obstante, en muchos otros tipos de investigaciones, es sumamente importante la naturaleza del participante. Por ejemplo, los alumnos universitarios probablemente no serían los participantes adecuados en estudios acerca de las actitudes hacia los niños, ya que la experiencia de los alumnos comúnmente no incluye la paternidad o maternidad. En el mismo sentido, no se puede analizar la capacidad de lectura en escuelas suburbanas y generalizar los resultados a todos los alumnos en todas las escuelas, o bien examinar la satisfacción laboral en la industria informática y generalizarla a todo tipo de industria.

Otro inconveniente es el modo en que se seleccionan los participantes de un estudio. Por ejemplo, en una encuesta por correspondencia acerca del conocimiento de un tema, algunos individuos devolverán el cuestionario y otros no. Presumiblemente existen diferencias sistemáticas entre aquellos que lo devuelven y aquellos que no, y es probable que aquellos que sí devuelven el cuestionario tengan más conocimientos acerca del tema en estudio. Si el investigador utiliza sólo los cuestionarios que fueron devueltos, podría llegar a la conclusión de que las personas tienen mayores conocimientos acerca de determinado tema que si hubiera podido analizar a toda la población. De modo similar, las personas que se ofrecen voluntariamente a participar en un experimento pueden diferir de aquellas que no lo hacen. Por ejemplo, los voluntarios pueden tener una personalidad más sensible a las necesidades ajenas.

El **muestreo aleatorio** es considerado el método óptimo para asegurar que una muestra sea representativa de su población. Muestreo aleatorio significa que los investigadores comienzan con una lista de todos los miembros de la población sobre la cual desean generalizar sus resultados (por ejemplo una lista de todos los psicoterapeutas de la nación), y luego utilizan un procedimiento al azar (tal como una tabla de números aleatorios) para seleccionar una muestra de esa población. El resultado del proceso descripto se denomina muestra probabilística, ya que cada miembro de la población estudiada tiene la misma probabilidad de ser incluido en la muestra del estudio.

No se debe confundir el muestreo aleatorio con la asignación aleatoria a los grupos que tratamos anteriormente. Ambos procesos utilizan verdaderos procedimientos al azar, pero el muestreo aleatorio se refiere al método de obtención de una muestra, y la asignación aleatoria se refiere al procedimiento de decisión con respecto a qué miembros de la muestra participarán en el grupo experimental y cuáles en el grupo de control.

MEDICIÓN

La cuarta condición mencionada como requisito para un estudio ideal es que las medidas deben ser precisas y adecuadas.

En la investigación psicológica existen tres tipos de medidas principales: **medidas de informe propio,** tales como cuestionarios o entrevistas; **medidas por observación o de comportamiento,** como por ejemplo las escalas de clasificación del comportamiento de niños mientras juegan, la cantidad de clientes que pasan por un molinete, la cantidad de milésimas de segundos para responder en un experimento que analiza el tiempo de reacción o la cantidad de veces que una rata presiona una barra, y, por último, **medidas fisiológicas**, como podrían ser los niveles hormonales o el ritmo cardíaco. Los tres tipos de mediciones se evalúan principalmente según su confiabilidad y validez.

Confiabilidad

La **confiabilidad** de una medida es su precisión o coherencia, es decir, en qué grado los resultados son similares si se aplica la misma medida al mismo elemento, en circunstancias idénticas. En psicología, los resultados no necesariamente son similares. Por ejemplo, cuestionarios entregados a las mismas personas en diferentes días dan con frecuencia resultados disímiles. A veces las preguntas son ambiguas y, por lo tanto, una persona puede responder de un modo en un momento y luego de otro. O bien, las personas pueden simplemente marcar en forma incorrecta alguna o todas las respuestas en una o más oportunidades. Las medidas de informe propio no son las únicas que pueden no resultar confiables. Las medidas por observación pueden no ser confiables debido a que los distintos observadores pueden estar en desacuerdo, y las medidas fisiológicas con frecuencia son sumamente erráticas entre un momento y otro.

Existen tres tipos de indicadores para medir el grado de confiabilidad: a) **la confiabilidad por prueba-reprueba**, conforme a la cual el mismo grupo es puesto a prueba dos veces; b) **la coherencia interna**, según la cual, por ejemplo, los puntos obtenidos en la mitad de las preguntas se comparan con los puntos obtenidos en la otra mitad (el alfa de Cronbach, descripta brevemente en el capítulo 17, es el método más común para determinar la coherencia interna), y c) **la confiabilidad por intercambio de juicios** utilizada para medidas de observación, es el grado de acuerdo entre los observadores. La tabla A-2 resume los tipos de confiabilidad descriptos.

Validez

La **validez** de una medida se refiere al hecho de que efectivamente pueda medir lo que pretende. (El término validez se aplica, asimismo, a estudios completos, cuando se refiere a lo apropiado de la conclusión que puede derivarse de los resultados).

Una medida que no es confiable no puede ser válida; una medida no confiable no mide nada. Pero aun cuando una medida sea confiable (precisa y repetible), no necesariamente es válida para medir lo que pretende medir. Por ejemplo, un cuestionario sobre satisfacción marital que pregunte, "¿cuál es la probabilidad de que usted permanezca con su esposo durante los próximos años? puede resultar sumamente confiable (por ejemplo, las personas pueden contestar las preguntas que incluye de forma bastante coherente), pero en lugar de medir satisfacción marital, podría estar midiendo el compromiso hacia el matrimonio; y los que responden el cuestionario podrían estar comprometidos no porque están satisfechos sino porque no tienen otra alternativa que la vida conyugal, o bien porque sienten que son muy poco atractivos y su situación sólo podría empeorar si abandonaran a su pareja.

Tabla A-2.
Tipos de confiabilidad.

Confiabilidad por prueba-reprueba:	correlación de pruebas aplicadas a las mismas personas en diferentes ocasiones.
Coherencia interna:	correlación entre los distintos ítems.
Confiabilidad por intercambio de juicios:	correlación entre los valores de diferentes evaluadores al calificar al mismo grupo de personas y objetos.

Otra razón por la cual una prueba puede no ser válida, aun siendo confiable, es que en lugar de medir la variable que se pretende medir, en realidad está midiendo una tendencia para intentar dar una buena impresión, o bien decir que sí o cualquier otro **sesgo de respuesta** por parte de los que responden. Una manera de encarar el problema de la intención de dar una buena impresión es incluir una "escala de deseo social", a veces llamada "escala de la mentira". Cuando la puntuación de un participante en una escala como la mencionada es alta, el investigador puede simplemente descartar la prueba realizada por el participante. Otra alternativa sería que los valores en una escala de deseo social puedan utilizarse en un procedimiento estadístico (tal como una correlación parcial o un análisis de covarianza, ambos descriptos brevemente en el capítulo 17) para adaptar el valor de esa persona en cuanto a la parte regular de la medida.

La validez de una medida es más difícil de evaluar que la confiabilidad. Para lograrlo se utilizan diversos medios. Existe **validez de contenido** cuando el contenido de la medida parece abarcar todos los distintos aspectos de aquello que se está midiendo. Usualmente, la validez de contenido la determina el investigador u otros expertos según el juicio de cada uno.

Asimismo, existen medios más sistemáticos para evaluar la validez de una medida. Determinar la **validez o criterio** implica realizar un estudio especial en el cual el investigador compara registros de la medida en cuestión con algún otro indicador posible de la misma variable. Por ejemplo, un investigador podría probar la validez de una medida de salud mental comparando valores de personas de un hospital psiquiátrico con las de puntuaciones de la población en general. Un tipo de validez de criterio es la **validez predictiva** de una medida. Por ejemplo, el hecho de que los registros de una prueba de capacidad laboral, tomada al presentarse la persona para solicitar un trabajo, predigan el desempeño efectivo de la persona en el empleo. La validez predictiva se utiliza especialmente cuando se diseña una medida con fines predictivos, como por ejemplo para la ubicación laboral o educativa. Otro tipo de validez de criterio es la **validez concurrente,** la cual se refiere al procedimiento de comparación de valores de una medida, con los de otra que mide directamente lo mismo; por ejemplo, una prueba de inteligencia nueva y breve comparada con una prueba de inteligencia existente más prolongada. La tabla A-3 resume los tres métodos de evaluación de la validez.

También puede aparecer el término **validez de constructo**, el cual se utiliza de varias formas (con frecuencia ambiguamente). Incluso textos sobre medición psicológica difieren en cuanto a este término. A veces incluyen la validez de criterio y, otras, la validez de contenido. Con frecuen-

Tabla A-3.
Tipos de validez de una medida.

Validez de contenido:	conforme la opinión de los expertos, el contenido de la prueba parece abarcar todo el espectro de lo que la prueba pretende medir.
Validez de criterio:	las puntuaciones de la prueba se correlacionan con algún otro indicador de lo que se supone mide la prueba.
Validez predictiva:	la puntuación de la prueba predice valores en otra variable que debería ser predicha por la prueba, conforme a lo que pretende medir; es un tipo de validez de criterio.
Validez concurrente:	la puntuación en la prueba se correlaciona con otra variable medida al mismo tiempo y que se sabe está relacionada con lo que la prueba pretende medir; es un tipo de validez de criterio.

cia, se utiliza para referirse a la medida que se utiliza en un estudio en el que existía un resultado predicho que fue confirmado por el estudio. Dado que la medida utilizada logró producir el resultado predicho, se demuestra que la idea (o "constructo") implícita en la medida queda comprobada conforme a la teoría.

Términos clave

- Medidas de comportamiento.
- Método de prueba a ciegas.
- Validez concurrente.
- Validez de contenido.
- Grupo control.
- Diseño de investigación correlacional.
- Compensación.
- Validez de criterio.
- Variable dependiente.
- Procedimiento doble ciego.
- Grupo experimental.
- Manipulación experimental.
- Sesgo del experimentador.
- Efectos del experimentador.
- Validez externa.
- Legitimidad de la generalización.
- Efectos Hawthorne.
- Variable independiente.
- Confiabilidad por intercambio de juicios.
- Coherencia interna.
- Validez interna.
- Manipulación de la variable independiente.
- Diseño de grupo de control equivalente con prueba previa y posterior.
- Diseño de investigación de grupo de control equivalente.
- Medidas por observación.
- Participantes.
- Medidas fisiológicas.
- Efectos placebo.
- Población.
- Validez predictiva.
- Diseño preexperimental.
- Muestra probabilística.
- Diseño cuasiexperimental.
- Asignación aleatoria a los grupos.
- Muestreo aleatorio.
- Confiabilidad.
- Diseño de investigación de medidas repetidas.
- Sesgo de respuesta.
- Muestra.
- Medidas de informe propio.
- Diseño de grupo único con prueba previa y posterior.
- Investigación de sujeto único.
- Confiabilidad por prueba-reprueba.
- Experimento verdadero.
- Validez.
- Diseño de investigación intrasujeto.

APÉNDICE

B Tablas

Tabla B-1.
Áreas bajo la curva normal:
Porcentaje del área bajo curva normal entre la media y las puntuaciones Z indicadas.

Z	% desde la media hasta la Z	Z	% desde la media hasta la Z	Z	% desde la media hasta la Z
0,00	0,00	0,24	9,48	0,48	18,44
0,01	0,40	0,25	9,87	0,49	18,79
0,02	0,80	0,26	10,26	0,50	19,15
0,03	1,20	0,27	10,64	0,51	19,50
0,04	1,60	0,28	11,03	0,52	19,85
0,05	1,99	0,29	11,41	0,53	20,19
0,06	2,39	0,30	11,79	0,54	20,54
0,07	2,79	0,31	12,17	0,55	20,88
0,08	3,19	0,32	12,55	0,56	21,23
0,09	3,59	0,33	12,93	0,57	21,57
0,10	3,98	0,34	13,31	0,58	21,90
0,11	4,38	0,35	13,68	0,59	22,24
0,12	4,78	0,36	14,06	0,60	22,57
0,13	5,17	0,37	14,43	0,61	22,91
0,14	5,57	0,38	14,80	0,62	23,24
0,15	5,96	0,39	15,17	0,63	23,57
0,16	6,36	0,40	15,54	0,64	23,89
0,17	6,75	0,41	15,91	0,65	24,22
0,18	7,14	0,42	16,28	0,66	24,54
0,19	7,53	0,43	16,64	0,67	24,86
0,20	7,93	0,44	17,00	0,68	25,17
0,21	8,32	0,45	17,36	0,69	25,49
0,22	8,71	0,46	17,72	0,70	25,80
0,23	9,10	0,47	18,08	0,71	26,11

Tabla B-1 (cont.).

Z	% desde la media hasta la Z	Z	% desde la media hasta la Z	Z	% desde la media hasta la Z
0,72	26,42	1,26	39,62	1,80	46,41
0,73	26,73	1,27	39,80	1,81	46,49
0,74	27,04	1,28	39,97	1,82	46,56
0,75	27,34	1,29	40,15	1,83	46,64
0,76	27,64	1,30	40,32	1,84	46,71
0,77	27,94	1,31	40,49	1,85	46,78
0,78	28,23	1,32	40,66	1,86	46,86
0,79	28,52	1,33	40,82	1,87	46,93
0,80	28,81	1,34	40,99	1,88	46,99
0,81	29,10	1,35	41,15	1,89	47,06
0,82	29,39	1,36	41,31	1,90	47,13
0,83	29,67	1,37	41,47	1,91	47,19
0,84	29,95	1,38	41,62	1,92	47,26
0,85	30,23	1,39	41,77	1,93	47,32
0,86	30,51	1,40	41,92	1,94	47,38
0,87	30,78	1,41	42,07	1,95	47,44
0,88	31,06	1,42	42,22	1,96	47,50
0,89	31,33	1,43	42,36	1,97	47,56
0,90	31,59	1,44	42,51	1,98	47,61
0,91	31,86	1,45	42,65	1,99	47,67
0,92	32,12	1,46	42,79	2,00	47,72
0,93	32,38	1,47	42,92	2,01	47,78
0,94	32,64	1,48	43,06	2,02	47,83
0,95	32,89	1,49	43,19	2,03	47,88
0,96	33,15	1,50	43,32	2,04	47,93
0,97	33,40	1,51	43,45	2,05	47,98
0,98	33,65	1,52	43,57	2,06	48,03
0,99	33,89	1,53	43,70	2,07	48,08
1,00	34,13	1,54	43,82	2,08	48,12
1,01	34,38	1,55	43,94	2,09	48,17
1,02	34,61	1,56	44,06	2,10	48,21
1,03	34,85	1,57	44,18	2,11	48,26
1,04	35,08	1,58	44,29	2,12	48,30
1,05	35,31	1,59	44,41	2,13	48,34
1,06	35,54	1,60	44,52	2,14	48,38
1,07	35,77	1,61	44,63	2,15	48,42
1,08	35,99	1,62	44,74	2,16	48,46
1,09	36,21	1,63	44,84	2,17	48,50
1,10	36,43	1,64	44,95	2,18	48,54
1,11	36,65	1,65	45,05	2,19	48,57
1,12	36,86	1,66	45,15	2,20	48,61
1,13	37,08	1,67	45,25	2,21	48,64
1,14	37,29	1,68	45,35	2,22	48,68
1,15	37,49	1,69	45,45	2,23	48,71
1,16	37,70	1,70	45,54	2,24	48,75
1,17	37,90	1,71	45,64	2,25	48,78
1,18	38,10	1,72	45,73	2,26	48,81
1,19	38,30	1,73	45,82	2,27	48,84
1,20	38,49	1,74	45,91	2,28	48,87
1,21	38,69	1,75	45,99	2,29	48,90
1,22	38,88	1,76	46,08	2,30	48,93
1,23	39,07	1,77	46,16	2,31	48,96
1,24	39,25	1,78	46,25	2,32	48,98
1,25	39,44	1,79	46,33	2,33	49,01

Tabla B-1 (cont.).

Z	% desde la media hasta la Z	Z	% desde la media hasta la Z	Z	% desde la media hasta la Z
2,34	49,04	2,58	49,51	2,82	49,76
2,35	49,06	2,59	49,52	2,83	49,77
2,36	49,09	2,60	49,53	2,84	49,77
2,37	49,11	2,61	49,55	2,85	49,78
2,38	49,13	2,62	49,56	2,86	49,79
2,39	49,16	2,63	49,57	2,87	49,79
2,40	49,18	2,64	49,59	2,88	49,80
2,41	49,20	2,65	49,60	2,89	49,81
2,42	49,22	2,66	49,61	2,90	49,81
2,43	49,25	2,67	49,62	2,91	49,82
2,44	49,27	2,68	49,63	2,92	49,82
2,45	49,29	2,69	49,64	2,93	49,83
2,46	49,31	2,70	49,65	2,94	49,84
2,47	49,32	2,71	49,66	2,95	49,84
2,48	49,34	2,72	49,67	2,96	49,85
2,49	49,36	2,73	49,68	2,97	49,85
2,50	49,38	2,74	49,69	2,98	49,86
2,51	49,40	2,75	49,70	2,99	49,86
2,52	49,41	2,76	49,71	3,00	49,87
2,53	49,43	2,77	49,72	3,50	49,98
2,54	49,45	2,78	49,73	4,00	50,00
2,55	49,46	2,79	49,74	4,50	50,00
2,56	49,48	2,80	49,74		
2,57	49,49	2,81	49,75		

Tabla B-2.
Puntos de corte para la distribución *t*.

	Pruebas de una cola			Pruebas de dos colas		
gl	*0,10*	*0,05*	*0,01*	*0,10*	*0,05*	*0,01*
1	3,078	6,314	31,821	6,314	12,706	63,657
2	1,886	2,920	6,965	2,920	4,303	9,925
3	1,638	2,353	4,541	2,353	3,182	5,841
4	1,533	2,132	3,747	2,132	2,776	4,604
5	1,476	2,015	3,365	2,015	2,571	4,032
6	1,440	1,943	3,143	1,943	2,447	3,708
7	1,415	1,895	2,998	1,895	2,365	3,500
8	1,397	1,860	2,897	1,860	2,306	3,356
9	1,383	1,833	2,822	1,833	2,262	3,250
10	1,372	1,813	2,764	1,813	2,228	3,170
11	1,364	1,796	2,718	1,796	2,201	3,106
12	1,356	1,783	2,681	1,783	2,179	3,055
13	1,350	1,771	2,651	1,771	2,161	3,013
14	1,345	1,762	2,625	1,762	2,145	2,977
15	1,341	1,753	2,603	1,753	2,132	2,947
16	1,337	1,746	2,584	1,746	2,120	2,921
17	1,334	1,740	2,567	1,740	2,110	2,898
18	1,331	1,734	2,553	1,734	2,101	2,879
19	1,328	1,729	2,540	1,729	2,093	2,861
20	1,326	1,725	2,528	1,725	2,086	2,846
21	1,323	1,721	2,518	1,721	2,080	2,832
22	1,321	1,717	2,509	1,717	2,074	2,819
23	1,320	1,714	2,500	1,714	2,069	2,808
24	1,318	1,711	2,492	1,711	2,064	2,797
25	1,317	1,708	2,485	1,708	2,060	2,788
26	1,315	1,706	2,479	1,706	2,056	2,779
27	1,314	1,704	2,473	1,704	2,052	2,771
28	1,313	1,701	2,467	1,701	2,049	2,764
29	1,312	1,699	2,462	1,699	2,045	2,757
30	1,311	1,698	2,458	1,698	2,043	2,750
35	1,306	1,690	2,438	1,690	2,030	2,724
40	1,303	1,684	2,424	1,684	2,021	2,705
45	1,301	1,680	2,412	1,680	2,014	2,690
50	1,299	1,676	2,404	1,676	2,009	2,678
55	1,297	1,673	2,396	1,673	2,004	2,668
60	1,296	1,671	2,390	1,671	2,001	2,661
65	1,295	1,669	2,385	1,669	1,997	2,654
70	1,294	1,667	2,381	1,667	1,995	2,648
75	1,293	1,666	2,377	1,666	1,992	2,643
80	1,292	1,664	2,374	1,664	1,990	2,639
85	1,292	1,663	2,371	1,663	1,989	2,635
90	1,291	1,662	2,369	1,662	1,987	2,632
95	1,291	1,661	2,366	1,661	1,986	2,629
100	1,290	1,660	2,364	1,660	1,984	2,626
∞	1,282	1,645	2,327	1,645	1,960	2,576

Tabla B-3.
Puntos de corte para la distribución *F*.

Denominador *gl*	Nivel de significación	Grados de libertad del numerador					
		1	*2*	*3*	*4*	*5*	*6*
1	0,01	4.052	5.000	5.404	5.625	5.764	5.859
	0,05	162	200	216	225	230	234
	0,10	39,9	49,5	53,6	55,8	57,2	58,2
2	0,01	98,50	99,00	99,17	99,25	99,30	99,33
	0,05	18,51	19,00	19,17	19,25	19,30	19,33
	0,10	8,53	9,00	9,16	9,24	9,29	9,33
3	0,01	34,12	30,82	29,46	28,71	28,24	27,91
	0,05	10,13	9,55	9,28	9,12	9,01	8,94
	0,10	5,54	5,46	5,39	5,34	5,31	5,28
4	0,01	21,20	18,00	16,70	15,98	15,52	15,21
	0,05	7,71	6,95	6,59	6,39	6,26	6,16
	0,10	4,55	4,33	4,19	4,11	4,05	4,01
5	0,01	16,26	13,27	12,06	11,39	10,97	10,67
	0,05	6,61	5,79	5,41	5,19	5,05	4,95
	0,10	4,06	3,78	3,62	3,52	3,45	3,41
6	0,01	13,75	10,93	9,78	9,15	8,75	8,47
	0,05	5,99	5,14	4,76	4,53	4,39	4,28
	0,10	3,78	3,46	3,29	3,18	3,11	3,06
7	0,01	12,25	9,55	8,45	7,85	7,46	7,19
	0,05	5,59	4,74	4,35	4,12	3,97	3,87
	0,10	3,59	3,26	3,08	2,96	2,88	2,83
8	0,01	11,26	8,65	7,59	7,01	6,63	6,37
	0,05	5,32	4,46	4,07	3,84	3,69	3,58
	0,10	3,46	3,11	2,92	2,81	2,73	2,67
9	0,01	10,56	8,02	6,99	6,42	6,06	5,80
	0,05	5,12	4,26	3,86	3,63	3,48	3,37
	0,10	3,36	3,01	2,81	2,69	2,61	2,55
10	0,01	10,05	7,56	6,55	6,00	5,64	5,39
	0,05	4,97	4,10	3,71	3,48	3,33	3,22
	0,10	3,29	2,93	2,73	2,61	2,52	2,46
11	0,01	9,65	7,21	6,22	5,67	5,32	5,07
	0,05	4,85	3,98	3,59	3,36	3,20	3,10
	0,10	3,23	2,86	2,66	2,54	2,45	2,39
12	0,01	9,33	6,93	5,95	5,41	5,07	4,82
	0,05	4,75	3,89	3,49	3,26	3,11	3,00
	0,10	3,18	2,81	2,61	2,48	2,40	2,33
13	0,01	9,07	6,70	5,74	5,21	4,86	4,62
	0,05	4,67	3,81	3,41	3,18	3,03	2,92
	0,10	3,14	2,76	2,56	2,43	2,35	2,28
14	0,01	8,86	6,52	5,56	5,04	4,70	4,46
	0,05	4,60	3,74	3,34	3,11	2,96	2,85
	0,10	3,10	2,73	2,52	2,40	2,31	2,24

Tabla B-3 (cont.).

Denominador *gl*	Nivel de significación	Grados de libertad del numerador					
		1	*2*	*3*	*4*	*5*	*6*
15	0,01	8,68	6,36	5,42	4,89	4,56	4,32
	0,05	4,54	3,68	3,29	3,06	2,90	2,79
	0,10	3,07	2,70	2,49	2,36	2,27	2,21
16	0,01	8,53	6,23	5,29	4,77	4,44	4,20
	0,05	4,49	3,63	3,24	3,01	2,85	2,74
	0,10	3,05	2,67	2,46	2,33	2,24	2,18
17	0,01	8,40	6,11	5,19	4,67	4,34	4,10
	0,05	4,45	3,59	3,20	2,97	2,81	2,70
	0,10	3,03	2,65	2,44	2,31	2,22	2,15
18	0,01	8,29	6,01	5,09	4,58	4,25	4,02
	0,05	4,41	3,56	3,16	2,93	2,77	2,66
	0,10	3,01	2,62	2,42	2,29	2,20	2,13
19	0,01	8,19	5,93	5,01	4,50	4,17	3,94
	0,05	4,38	3,52	3,13	2,90	2,74	2,63
	0,10	2,99	2,61	2,40	2,27	2,18	2,11
20	0,01	8,10	5,85	4,94	4,43	4,10	3,87
	0,05	4,35	3,49	3,10	2,87	2,71	2,60
	0,10	2,98	2,59	2,38	2,25	2,16	2,09
21	0,01	8,02	5,78	4,88	4,37	4,04	3,81
	0,05	4,33	3,47	3,07	2,84	2,69	2,57
	0,10	2,96	2,58	2,37	2,23	2,14	2,08
22	0,01	7,95	5,72	4,82	4,31	3,99	3,76
	0,05	4,30	3,44	3,05	2,82	2,66	2,55
	0,10	2,95	2,56	2,35	2,22	2,13	2,06
23	0,01	7,88	5,66	4,77	4,26	3,94	3,71
	0,05	4,28	3,42	3,03	2,80	2,64	2,53
	0,10	2,94	2,55	2,34	2,21	2,12	2,05
24	0,01	7,82	5,61	4,72	4,22	3,90	3,67
	0,05	4,26	3,40	3,01	2,78	2,62	2,51
	0,10	2,93	2,54	2,33	2,20	2,10	2,04
25	0,01	7,77	5,57	4,68	4,18	3,86	3,63
	0,05	4,24	3,39	2,99	2,76	2,60	2,49
	0,10	2,92	2,53	2,32	2,19	2,09	2,03
26	0,01	7,72	5,53	4,64	4,14	3,82	3,59
	0,05	4,23	3,37	2,98	2,74	2,59	2,48
	0,10	2,91	2,52	2,31	2,18	2,08	2,01
27	0,01	7,68	5,49	4,60	4,11	3,79	3,56
	0,05	4,21	3,36	2,96	2,73	2,57	2,46
	0,10	2,90	2,51	2,30	2,17	2,07	2,01
28	0,01	7,64	5,45	4,57	4,08	3,75	3,53
	0,05	4,20	3,34	2,95	2,72	2,56	2,45
	0,10	2,89	2,50	2,29	2,16	2,07	2,00

Tabla B-3 (cont.).

Denominador *gl*	Nivel de significación	Grados de libertad del numerador					
		1	*2*	*3*	*4*	*5*	*6*
95	0,01	6,91	4,84	4,00	3,52	3,22	3,00
	0,05	3,94	3,09	2,70	2,47	2,31	2,20
	0,10	2,76	2,36	2,14	2,01	1,91	1,84
100	0,01	6,90	4,82	3,98	3,51	3,21	2,99
	0,05	3,94	3,09	2,70	2,46	2,31	2,19
	0,10	2,76	2,36	2,14	2,00	1,91	1,83
∞	0,01	6,64	4,61	3,78	3,32	3,02	2,80
	0,05	3,84	3,00	2,61	2,37	2,22	2,10
	0,10	2,71	2,30	2,08	1,95	1,85	1,78

Tabla B-4.
Puntos de corte para la distribución chi-cuadrado.

	Nivel de significación		
gl	*0,10*	*0,05*	*0,01*
1	2,706	3,841	6,635
2	4,605	5,992	9,211
3	6,252	7,815	11,345
4	7,780	9,488	13,277
5	9,237	11,071	15,087
6	10,645	12,592	16,812
7	12,017	14,067	18,475
8	13,362	15,507	20,090
9	14,684	16,919	21,666
10	15,987	18,307	23,209

Tabla B-5.
Índice de las tablas de potencia y de las tablas con la cantidad de participantes necesarios para obtener una potencia del 80%.

Procedimiento de prueba de hipótesis	Capítulo	Tabla de potencia	Tabla con cantidad de participantes
Coeficiente de correlación (r)	3	99	99
Prueba t para medias dependientes	9	278	279
Prueba t para medias independientes	10	308	309
Análisis de varianza de un criterio	11	341	341
Análisis de varianza de dos criterios	13	410	410
Prueba de chi-cuadrado de independencia	14	450	450

Respuestas a los ejercicios de la serie I

Capítulo 1

1. a) Nominal (o categórico).
   b) Numérico (o cuantitativo); más precisamente, intervalar.
   c) Numérico (o cuantitativo); más precisamente, ordinal.
2. a) Tabla de frecuencias.

Valores	Frecuencia	Valores	Frecuencia
96	1	72	0
95	0	71	1
94	0	70	1
93	0	69	1
92	1	68	2
91	1	67	1
90	0	66	0
89	0	65	0
88	0	64	2
87	1	63	0
86	0	62	0
85	1	61	0
84	0	60	0
83	2	59	1
82	0	58	0
81	1	57	0
80	1	56	0
79	0	55	0
78	0	54	0
77	0	53	0
76	2	52	0
75	2	51	0
74	1	50	1
73	1		

b) Tabla de frecuencias agrupadas (una de varias posibilidades).

Intervalo	Frecuencia
95–99	1
90–94	2
85–89	2
80–84	4
75–79	4
70–74	4
65–69	4
60–64	2
55–59	1
50–54	1

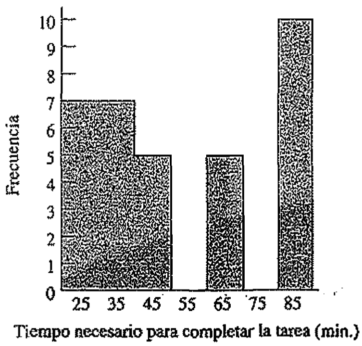
c) Histograma (según tabla del punto b).

d) Forma general de la distribución: unimodal, aproximadamente simétrica (leve asimetría negativa).

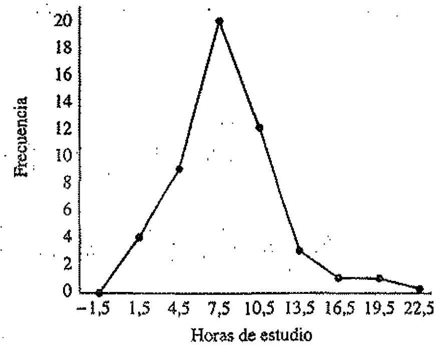
3. a) Tabla de frecuencias agrupadas (una de varias posibilidades).

Intervalo	Frecuencia
80–89	10
70–79	0
60–69	5
50–59	0
40–49	5
30–39	7
20–29	7

b) Histograma (según tabla del punto a).



c) Polígono de frecuencias (según tabla del punto b).



d) Forma general de la distribución: unimodal, asimétrica hacia la derecha (positivamente asimétrica).

4. a) Tabla de frecuencia.

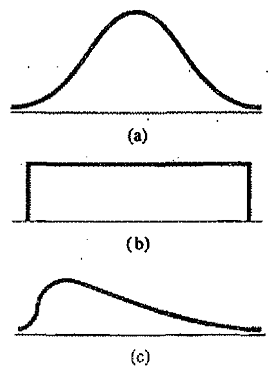
Cantidad de horas	Frecuencia	Cantidad de horas	Frecuencia
18	1	8	5
17	0	7	11
16	0	6	4
15	1	5	2
14	0	4	3
13	2	3	4
12	1	2	2
11	3	1	1
10	5	0	1
9	4		

b) Tabla de frecuencias agrupadas (una de varias posibilidades).

Intervalo	Frecuencia
18–20	1
15–17	1
12–14	3
9–11	12
6–8	20
3–5	9
0–2	4

5. a) Bimodal; b) aproximadamente normal (o unimodal o simétrica); c) multimodal.

6.



7. a) Una distribución es la forma en que un grupo de valores se organiza entre los diferentes valores posibles. Una manera de describir tal distribución es a través de un gráfico, denominado histograma. Un histograma es un tipo de gráfico de barras con una barra para cada valor posible, ordenadas de izquierda a derecha. Las barras tienen una altura igual a la cantidad de veces según el valor que representan es observado. En este tipo de gráficos, una distribución simétrica tiene forma simétrica, es decir que la mitad derecha y la mitad izquierda semejan imágenes especulares. En un sentido amplio, significa que existen prácticamente la

misma cantidad de valores altos que bajos, y que a medida que nos movemos del valor medio hacia el valor más alto o el más bajo, la cantidad de observaciones de cada valor disminuye o aumenta del mismo modo).
Una distribución es unimodal si el histograma tiene un punto alto. Es decir que existe un sólo nivel en particular que presenta más frecuencias que cualquier otro nivel. (A este nivel se lo denomina "moda", y ser unimodal significa tener sólo una moda) .

b) Una distribución unimodal negativamente asimétrica no es simétrica, y su cola, el lado bajo y alargado de la misma, se extiende hacia la izquierda (adonde se encuentran los valores negativos del gráfico).

Capítulo 2

1. Serie A. a) $M = \Sigma X/N = 261/9 = 29$.
   b) Mediana = 28.
   c) $SS = \Sigma(X-M)^2 = (32-29)^2 + (28-29)^2 + (24-29)^2 + (28-29)^2 + (28-29)^2 + (31-29)^2 + (35-29)^2 + (29-29)^2 + (26-29)^2$
   $SS = 3^2 + (-1)^2 + (-5)^2 + (-1)^2 + (-1)^2 + 2^2 + 6^2 + 0^2 + (-3)^2$
   $SS = 9+1+25+1+1+4+36+0+9 = 86$.
   d) $SD^2 = SS/N = \Sigma(X-M)^2/N = 86/9 = 9{,}56$.
   e) $SD = \sqrt{SD^2} = \sqrt{9{,}56} = 3{,}09$.
   Serie B. a) $M = 4$; b) Mediana = 4; c) $SS = 26$; d) $SD^2 = 3{,}25$; e) $SD = 1{,}80$.

2. La temperatura promedio, entendiendo como tal la suma de las 10 lecturas dividida por 10, fue −7 grados Celsius. Este resultado es la **media**. Sin embargo, si ordenamos las temperaturas de menor a mayor, los dos números del medio que determinan lo que se denomina la temperatura **mediana**, serían igual a −5 grados. Otra forma de representar la temperatura típica sería tomar la temperatura específica que ocurrió más frecuentemente, a la que se llama **moda**. En este caso, hubo dos modas, dos temperaturas que ocurrieron más frecuentemente, −1 y −5. Las dos temperaturas ocurrieron dos veces. Pero la moda no es una información muy útil en este caso.
   Con respecto a la variación, se puede calcular según la medida en que varió cada temperatura con respecto al promedio: primero, se eleva al cuadrado cada uno de esos "desvíos" (con este procedimiento anulamos los signos positivos y negativos de manera que obtenemos la diferencia con respecto al promedio sin importar el sentido de la misma). Luego calculamos el promedio de estos desvíos cuadráticos. Por ejemplo, el desvío de la primera temperatura es 2 (es decir, −5 menos −7) que, elevado al cuadrado, da 4. Elevando cada desvío al cuadrado y sumándolos obtenemos un resultado de 468. Al dividir este resultado por 10 obtenemos un desvío cuadrático medio de 46,8. A este resultado se lo denomina varianza. La varianza es una forma de describir la dispersión de un grupo de números. La varianza es una parte muy importante de muchos cálculos estadísticos. Sin embargo, lamentablemente no transmite una idea muy directa del grado en que varían los números.
   Podemos obtener una idea más directa del grado de variación de un grupo de números entre sí calculando la raíz cuadrada de la varianza. En este caso, la raíz cuadrada de 46,8 es 6,84. La raíz cuadrada de la varianza se denomina **desvío estándar**. Sin entrar en detalles, significa que, en un día promedio, la temperatura difiere en 6,84 grados con respecto al promedio de −7 grados.

3. El resultado consta de dos partes. En primer lugar, la "media" se refiere al promedio aritmético común: sumar la cantidad total de sueños y dividirlos por la cantidad de personas. En este caso, la cantidad promedio de sueños narrados durante las dos semanas fue 6,84. En segundo lugar, el "*SD*" se refiere al "desvío estándar". El desvío estándar es, en un sentido amplio, el promedio de dispersión de la cantidad de sueños con respecto al promedio de dichos sueños; en este caso la dispersión es de 3,18 sueños. La dispersión es bastante amplia. Para ser más precisos, el desvío estándar se calcula tomando la cantidad de sueños de cada persona, restando 6,84 a esas cantidades y elevando la diferencia al cuadrado; el desvío estándar es la raíz cuadrada del promedio de esas diferencias cuadráticas.

4. a) $Z = (X-M)/SD = (91-79)/12 = 12/12 = 1{,}00$.
   b) $Z = (68-79)/12 = -11/12 = -0{,}92$
   c) $Z = (103-79)/12 = 24/12 = 2{,}00$.

5. a) Si el CI = 107, $Z = (X-M)/SD = (107-100)/16 = 7/16 = 0{,}44$.
   $X = (Z)(SD) + M = (0{,}44)(41) + 231 = 18{,}04 + 231 = 249$.
   (El resultado final está redondeado a un número entero ya que el valor real de la prueba se refiere a la cantidad de puntuaciones correctas, lo cual no puede ser una fracción)
   b) $Z = -1{,}06$; $X = 188$.
   c) $Z = 0$; $X = 231$.

6. Esposa: $Z = (X-M)/SD = (63-60)/6 = 3/6 = 0{,}5$.
   Esposo: $Z = (X-M)/SD = (59-55)/4 = 4/4 = 1$.
   El esposo presenta una puntuación Z más elevada, de lo que se deduce que se ha adaptado mejor, con relación a otros hombres divorciados, de lo que lo ha hecho su esposa con relación a otras mujeres divorciadas.

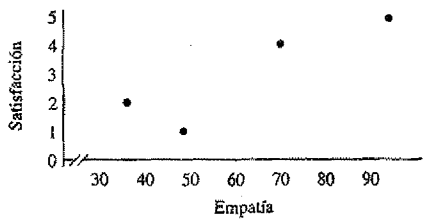
Explicación a una persona que nunca ha asistido a un curso de estadística: para el caso de las esposas, un valor observado de 63 es 3 puntos mejor que el promedio 60 para mujeres divorciadas en general. (La "media" del problema es un término estadístico con el que se denomina el promedio ordinario, la suma de los valores dividida por la cantidad de valores). Pero existen por supuesto algunas variaciones entre los valores observados de mujeres divorciadas. La cantidad promedio aproximada en que los valores de las mujeres difieren del promedio es 6 puntos; es el *SD* al que se refiere el problema. (En realidad, y sin entrar en detalles, el *SD*, que significa desvío estándar, no es más que el promedio de las desviaciones de los valores observados con respecto al promedio. Para ser más precisos, el *SD* es la raíz cuadrada del promedio de las diferencias cuadráticas de cada valor con la media).

Por lo tanto, el valor de la esposa supera la media de las esposas, en general, en una cantidad que es sólo la mitad de lo que los valores de las esposas en general varían de la media correspondiente. Esto le da lo que se denomina una puntuación Z de +0,5, que la ubica en una escala en la que se compara su valor con el de las mujeres divorciadas en general. Utilizando la misma lógica para analizar la adaptación al divorcio por parte del marido, en comparación con otros hombres divorciados, él se encuentra por sobre el promedio en una cantidad igual al promedio según el cual los hombres varían de la media, es decir, presenta una puntuación Z de +1.

Por lo tanto, aunque los dos se han adaptado mejor que el promedio para su sexo, el esposo se adaptó mejor, en relación con otros hombres divorciados, de lo que la esposa se adaptó con relación a otras mujeres divorciadas.

Capítulo 3

1. a)



b) Correlación lineal positiva. A medida que aumenta la empatía del terapeuta también aumenta la satisfacción del paciente.

c)

	Empatía terapeuta		Satisfacción paciente		$Z_X Z_Y$
	Original	Z_X	*Original*	Z_Y	
1	70	0,36	4	0,63	0,23
2	94	1,45	5	1,26	1,83
3	36	−1,17	2	−0,63	0,74
4	48	−0,63	1	−1,26	0,80
					$\Sigma = 3,60$

$r = 3,60/4 = 0,90$

d) El primer paso para resolver un ejercicio de correlación es realizar un gráfico, representar una variable en cada eje y después marcar un punto en el lugar correspondiente a cada observación en ese gráfico. A esto se lo denomina diagrama de dispersión, y nos da una imagen del patrón de relación entre las dos variables.

En este caso, los valores altos parecen coincidir con los altos, y los bajos con los bajos, determinando lo que se denomina una **correlación positiva**. (Básicamente, la correlación indica en qué medida los valores altos coinciden con los valores altos y los valores bajos con los valores bajos). Además, dado que los puntos se ubican aproximadamente cerca de una línea recta, podemos decir que es un ejemplo de **correlación lineal positiva**.

El siguiente paso es convertir todos los valores observados en puntuaciones Z para facilitar el cálculo del grado en el que los altos coinciden con los altos y los bajos con los bajos. Las puntuaciones Z facilitan este cálculo porque constituyen el mejor indicador de cuán bajo o alto es un valor en relación con los otros valores de la distribución.

El coeficiente de correlación es un número que indica el grado de asociación. Se calcula multiplicando las dos puntuaciones Z de cada persona, sumando estos productos y luego promediando el total por la cantidad de personas involucradas en el estudio. El coeficiente será un número alto si los registros altos coinciden con los altos y los bajos con los bajos, debido a que en el caso de las puntuaciones Z, los altos son siempre positivos (y cuanto más altos son, más positivos) y al multiplicar positivo por positivo el resultado es positivo. Por otro lado, con respecto a las puntuaciones Z, los bajos son siempre negativos (y cuanto más bajo el registro, más negativa es la puntuación

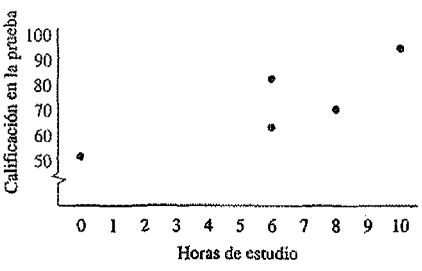
Z), y al multiplicar negativo por negativo el resultado también es positivo.

Los estadísticos pueden probar que, siguiendo con este procedimiento, el número más alto que se puede obtener, si los valores de las dos variables están perfectamente correlacionados, es +1. Si no existiera relación lineal entre las variables, el resultado de este procedimiento sería 0 (obtendríamos 0 porque los valores altos serían multiplicados a veces por altos y a veces por bajos, y los valores bajos a veces por altos y a veces por bajos, dando una mezcla de números positivos y negativos que se cancelarían entre sí).

En el caso que estamos analizando aquí, el total de los productos de las puntuaciones Z es 3,6, y al dividirlo por la cantidad de parejas terapeuta-paciente da 0,90. A este resultado, 0,9, se lo denomina **coeficiente *r* de correlación de Pearson,** e indica una fuerte correlación lineal positiva entre la satisfacción y la empatía.

e) Existen tres posibilidades lógicas de la dirección de causalidad: (i) Si un terapeuta tiene más empatía, esto hace que el paciente se sienta más satisfecho (la empatía causa satisfacción); (ii) si un paciente se siente más satisfecho, esto puede hacer que el terapeuta sienta más empatía hacia el paciente (la satisfacción causa empatía), o (iii) algún tercer factor, como una buena adaptación entre el problema del paciente y la habilidad del terapeuta, puede hacer que los pacientes estén más satisfechos y que los terapeutas sientan más empatía (un tercer factor causa tanto la satisfacción como la empatía).

2. a)



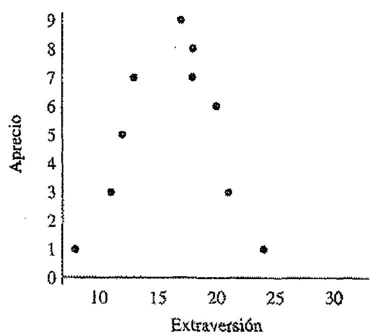
b) Correlación lineal positiva. A medida que aumentan las horas de estudio también lo hacen las calificaciones.

Horas de estudio $Z_X Z_Y$		**Calificaciones en la prueba**		
Original	Z_X	*Original*	Z_Y	
0	–1,79	52	–1,41	2,52
10	1,19	95	1,48	1,76
6	0,00	83	0,67	0,00
8	0,60	71	–0,13	–0,08
6	0,00	64	–0,60	0,00
$M = 6$;	$SD = 3{,}35$	$M = 73$;	$SD = 14{,}90$	$\Sigma = 4{,}20$
		$r = 4{,}20/5 = 0{,}84$.		

d) *Véase* en la respuesta al ejercicio 1d un ejemplo de cómo escribir un ensayo de este tipo.

e) Existen tres direcciones de causalidad lógicas posibles:
(i) Estudiar más horas es la causa de mejores calificaciones; (ii) obtener mejores calificaciones es la causa de más horas de estudio. Cabe destacar que, aunque en la teoría esto sea posible, en la realidad es imposible que un hecho futuro (la calificación en la prueba) cause un evento anterior (horas de estudio), o (iii) un tercer factor, como por ejemplo el interés por la materia, podría ser la causa de que el alumno estudie más y de que le vaya mejor en la prueba.

3. a)



b) Correlación curvilínea (forma de U invertida): hasta determinado punto, a medida que aumenta la extraversión, aumenta el aprecio, pero superado ese punto, a medida que la extraversión continúa aumentando el aprecio disminuye.

c)

Extraversión de un integrante		Aprecio por ese integrante		$Z_X Z_Y$
Valor observado	*Puntuación Z*	*Valor observado*	*Puntuación Z*	
18	0,37	8	1,10	0,407
17	0,17	9	1,47	0,245
20	0,80	6	0,37	0,296
8	–1,72	1	–1,47	2,528
13	–0,67	7	0,74	–0,496
24	1,63	1	–1,47	–2,396
11	–1,09	3	–0,74	0,807
12	–0,88	5	0,00	0,000
18	0,38	7	0,74	0,281
21	1,00	3	–0,74	–0,740
				$\Sigma = 0{,}932$

$r = 0{,}932/10 = 0{,}09.$

4. a) La tabla muestra el grado de asociación entre los valores de varias medidas aplicadas a mujeres embarazadas y a sus compañeros. El grado de asociación entre valores de dos medidas indica en qué grado los valores altos o bajos en una medida coinciden con los valores altos o bajos en la otra medida. Los números que indican el grado de asociación se denominan **coeficientes de correlación.** Un coeficiente de correlación 1 significaría que los valores de las dos medidas estaban perfectamente vinculados; conocer el valor de una medida es todo lo que necesitaríamos para poder calcular el valor en la otra medida. (Estas asociaciones tan perfectas casi nunca ocurren en la vida real). Un coeficiente de correlación 0 significa que no existe asociación entre las dos medidas, los valores de una medida no tienen ninguna relación con los valores de la otra medida. Finalmente, un coeficiente de correlación –1 significa que existe una asociación inversa perfecta, es decir, valores altos en una medida están perfectamente asociados con los bajos en la otra, y viceversa. La mayoría de las correlaciones se encuentran entre 0 y +1 ó 0 y –1. Cuanto más cerca de 0 se encuentra una correlación, más débil es el grado de asociación.

   Por ejemplo, la correlación 0,17, entre lo informado por las mujeres sobre su estrés y lo informado por los hombres sobre el estrés de las mujeres, indica que la asociación entre estas dos medidas es bastante débil. Por lo tanto, el nivel de estrés que sufre una mujer no está muy relacionado con el nivel de estrés que su compañero creé que ella está sufriendo. Por otro lado, la correlación de 0,50 (cerca del medio de la primera columna de correlaciones) indica que existe una asociación mucho más fuerte entre el nivel de estrés informado por una mujer y su estado depresivo al momento de la segunda entrevista. Es decir, es probable que las mujeres que informan estar bajo cierto nivel de estrés también informen estar deprimidas; aquellas que no están bajo mucho estrés probablemente no informen estar muy deprimidas.

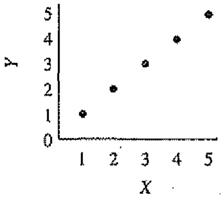
   b) En general, de todas las correlaciones representadas en esta tabla, las más fuertes se dan entre las variables estrés, apoyo y estado emocional; las correlaciones de estas variables con las demográficas (edad, origen étnico, etc.) eran bastante débiles. El apoyo del compañero parecía estar fuertemente correlacionado con el estrés y el estado anímico, y el estado depresivo al momento de la segunda prueba estaba particularmente relacionado con las otras variables.

   c) Sólo porque dos variables están correlacionadas, aun cuando estén fuertemente correlacionadas no significa que podamos conocer la dirección de causalidad particular que crea dicha asociación. Por ejemplo, existe una fuerte correlación inversa entre el apoyo del compañero en la primera prueba y el estado depresivo en la segunda. Existen tres direcciones de causalidad lógicamente posibles en este caso: el apoyo por parte del compañero puede causar una menor depresión; una menor depresión puede causar un mayor apoyo; o algún tercer factor puede causar ambos. Podemos anular la segunda posibilidad, ya que un hecho futuro (poca depresión) no puede causar un hecho pasado (apoyo inicial). Sin embargo, las otras dos posibilidades permanecen. Es realmente posible que el hecho de tener el apoyo del compañero ayude a reducir la depresión. Pero también es posible que un tercer factor esté causando ambas cosas. Por ejemplo, consideremos el nivel de ingresos. Tal vez cuando una pareja logra tener mayores ingresos, el compañero tiene más tiempo y energía para brindar su apoyo, y a la vez una mayor calidad de vida mantiene bajos los niveles de depresión.

5. a) Ambas medidas pueden presentar un bajo nivel de confiabilidad, reduciendo (atenuando) así la posible correlación entre ellas.

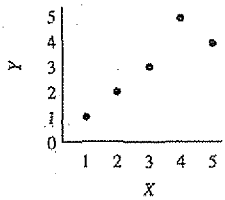
   b) Entre millonarios no puede haber un gran rango de calidad de vida (probablemente todos ellos tienen una muy buena calidad de vida), por lo tanto, la correlación con cualquier variable (incluso la variable felicidad) es limitada.

6. **SERIE A:**



X		Y		Producto cruzado de puntuaciones Z
Original	*Z*	*Original*	*Z*	
1	–1,41	1	–1,41	2,0
2	–0,71	2	–0,71	0,5
3	0,00	3	0,00	0,0
4	0,71	4	0,71	0,5
5	1,41	5	1,41	2,0
$M = 3$; $SD = 1,41$				5,0
				$r = 5,0/5 = 1,00$.

SERIE B :



X		Y		Producto cruzado de puntuaciones Z
Original	*Z*	*Original*	*Z*	
1	–141	1	–1,41	2,0
2	–0,71	2	–0,71	0,5
3	0,00	3	0,00	0,0
4	0,71	5	1,41	1,0
5	1,41	4	0,71	1,0
				4,5
				r = 4,5/5 = 0,90.

7. **SERIE A:**

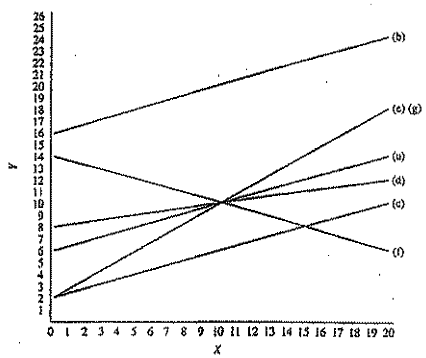
Toma medicamento		Se resfría		Producto cruzado de puntuaciones Z
Original	*Z*	*Original*	*Z*	
0	–1	1	1	–1
0	–1	1	1	–1
0	–1	1	1	–1
0	–1	1	1	–1
1	1	0	–1	–1
1	1	0	–1	–1
1	1	0	–1	–1
1	1	0	–1	–1
				–8
				$r = –8/8 = –1,00$.

SERIE B:

Toma medicamento		Se resfría		Producto cruzado de puntuaciones Z
Original	*Z*	*Original*	*Z*	
0	–1	1	1	–1
0	–1	1	1	–1
0	–1	1	1	–1
0	–1	0	–1	1
1	1	1	1	1
1	1	0	–1	–1
1	1	0	–1	–1
1	1	0	–1	–1
				–4
				$r = –4/8 = –0,50$.

Capítulo 4

1. a) Variable Predictora = puntuación en la prueba de conocimientos sobre fisiología. Variable dependiente = cantidad de lesiones durante el año subsiguiente. Beta = 0,4 (el coeficiente de correlación).
   b) $(Z)_{Lesiones} = (0,4)(Z_{fisiología})$.
   c) $(0,4)(-2) = -0,8$; $(0,4)(-1)$
   $= -0,4$; $(0,4)(0) = 0$; $(0,4)(1)$
   $= 0,4$; $(0,4)(2) = 0,8$.
2. a) $b = (\beta)(SD_Y/SD_X) = (0,4)(2/2) = 0,4$;
   $a = M_Y - (b)(M_X) = 10 - (0,4)(10)$
   $= 10 - 4 = 6$; $\hat{Y} = 6 + (0,4)(X)$.
   b) $Y = 16 + (0,4)(X)$.
   c) $Y = 2 + (0,4)(X)$.
   d) $Y = 8 + (0,2)(X)$.
   e) $Y = 2 + (0,8)(X)$.
   f) $Y = 14 + (-0,4)(X)$.
   $Y = 2 + (0,8)(X)$.



3.

Nota en el parcial	Modelo de predicción	Nota predicha en el final
30	40 + (0,5)(30)	55
40	40 + (0,5)(40)	60
50	40 + (0,5)(50)	65
60	40 + (0,5)(60)	70
70	40 + (0,5)(70)	75
80	40 + (0,5)(80)	80
90	40 + (0,5)(90)	85
100	40 + (0,5)(100)	90

4. a) $b = (b)(SD_Y/SD_X) = (0{,}9)(1{,}58/22{,}14)$
$= 0{,}064$;
$a = M_Y - (b)(M_X) = 3 - (0{,}064)(62)$
$= -0{,}97$;
Satisfacción predicha = –0,97 + (0,064) (empatía)

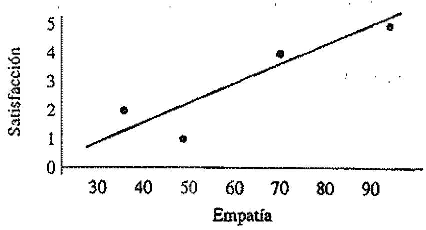
b)

Número de pareja	Empatía terapeuta	Satisfacción del cliente	
		Real	*Predicha*
1	70	4	3,51
2	94	5	5,05
3	36	2	1,33
4	48	1	2,10

c)

Error	Error^2
0,49	0,24
–0,05	0,00
0,67	0,45
–1,10	1,21

d)



e) Reducción proporcional de error
$= (SS_{Total} - SS_{Error})/SS_{Total}$
$= (10 - 1{,}9)/10 = 0{,}81$.

f) $\sqrt{0{,}81} = 0{,}9$; $r = 0{,}9$.

g) Puede comprobarse matemáticamente que el método más preciso para predecir la puntuación *Z* de una persona en una variable (llamémosla *Y*), sobre la base de la puntuación *Z* de esa persona en otra variable (llamémosla *X*), es multiplicar la puntuación *Z* en *X* por el coeficiente de correlación. Este procedimiento puede simplificarse en una sola fórmula (que no requiere las conversiones a puntuaciones *Z* y de puntuaciones *Z*), en la que las puntuaciones originales de *Y* pueden predecirse directamente a partir de las puntuaciones originales de *X*. En este ejemplo en particular, la fórmula es tal que, para predecir la puntuación original de un paciente en cuanto a satisfacción, se toma la constante de –0,97 y luego se le suma el resultado de multiplicar 0,064 por el valor correspondiente a la empatía del terapeuta.

Para evaluar la precisión de la fórmula se deben seguir los siguientes pasos. Primero, determinar la predicción que se hubiera hecho utilizando la fórmula para cada paciente de las cuatro parejas utilizadas para el cálculo del coeficiente de correlación. Por ejemplo, aplicando esta fórmula a la primera pareja, sumamos a –0,97 el resultado de multiplicar 0,064 por la empatía del terapeuta (0,064 x 70 = 4,48); el resultado es 3,51. Se puede calcular el error en el que incurriríamos utilizando este modelo para cada una de las predicciones, restando el valor predicho al valor observado. Por ejemplo, en el caso de la primera pareja, 4 menos 3,51 da un error de 0,49. Dado que los errores se cancelarían unos a otros al sumarlos (porque algunos son negativos y otros positivos), elevo los errores al cuadrado. Para ilustrarlo gráficamente, se trazó sobre el diagrama de dispersión preparado para estos datos una recta (denominada **recta de** regresión) que muestra las predicciones logradas utilizando la fórmula. Como se observa, los puntos correspondientes a los valores observados están bastante cerca de la recta de regresión, la distancia entre cada punto y la recta es el error.

Después se compara el **error** en el que incurriríamos, utilizando la fórmula de predicción, con el error en el que incurriríamos prediciendo sin ella (predecir sin ella significa predecir sólo con la media de los valores correspondientes a la satisfacción del paciente). El cálculo estadístico que en realidad se utiliza se denomina **reducción proporcional de error.** Es la reducción del error cuadrático al utilizar la fórmula (es decir, el error cuadrático total al predecir utilizando la media, que se calculó en 10, menos la suma de errores cuadráticos utilizando la fórmula, que se calculó en 1,9), dividido por el error cuadrático total al utilizar la media para predecir. Esto arroja un resultado de 0,81, lo que significa que el error cuadrático se ha reducido en un 81% con respecto a utilizar sólo la media para predecir. Dado que la reducción proporcional de error es matemáticamente equivalente al coeficiente de correlación cuadrático, el resultado se controló calculando la raíz cuadrada de la reducción proporcional de error. La raíz cuadrada de 0,81 es 0,9, que coincide exactamente con el coeficiente de correlación.

5. a) $b = (\beta)(SD_Y/SD_X)$
$= (0,84)(14,9/3,35) = 3,74$;
$a = M_Y - (b)(M_X) = 73 - (3,74)(6) = 50,56$.
Calificación predicha = 50,56 + (3,74) (horas de estudio).

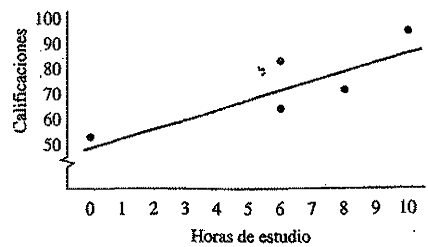
b)

Horas de estudio	Calificaciones (Y)	
(X)	*Original*	*Predicción*
0	52	50,56
10	95	87,96
6	83	73,00
8	71	80,48
6	64	73,00

c)

Error	Error^2
1,44	2,07
7,07	49,56
10,00	100,00
−9,48	89,87
−9,00	81,00

d)



e) Reducción proporcional de error
$= (SS_{Total} - SS_{Error})/SS_{Total} =$
$= (1.110 - 322,5)/1.110 = 0,71$.

f) $\sqrt{0,71} = 0,84$; $r = 0,84$.

g) Respuesta similar a la del ejercicio 4g.

6. Para realizar el estudio se utilizó un procedimiento estadístico denominado regresión múltiple. Este procedimiento produce una fórmula para predecir el valor de una persona en una variable dependiente (en este caso, la aceptación del niño por parte de sus pares) a partir de los registros de esa persona en una serie de variables de predicción (en este caso, la enseñanza no social y el entrenamiento social por parte de la madre del niño). La fórmula o ecuación se forma multiplicando el valor observado de esa persona, en cada una de las variables de predicción, por un número particular denominado coeficiente de regresión, y sumando luego los productos. Este procedimiento produce la regla de predicción más precisa en su tipo.

Cuando el coeficiente de regresión es el que se utiliza con puntuaciones Z, se denomina coeficiente de regresión estandarizado y se simboliza con la letra griega beta (β). En este ejemplo, estamos interesados en la ecuación 1, relacionada con la aceptación por parte de los pares. La tabla muestra los coeficientes beta para esta ecuación (0,10 y 0,32). Por lo tanto, la puntuación Z predicha para la aceptación por parte de los pares de un niño es 0,10 por la puntuación Z correspondiente a la enseñanza no social brindada por su madre, más 0,32 por la puntuación Z correspondiente al entrenamiento social brindado por su madre.

Los coeficientes de regresión sugieren que la aceptación de un niño por parte de sus pares está muy fuertemente relacionada con el entrenamiento social brindado por su madre, y mucho menos fuertemente relacionada con la enseñanza no social brindada por ella. Es importante destacar, sin embargo, que los coeficientes de regresión para cada una de estas variables de predicción reflejan lo que cada una contribuye a la predicción, independientemente de lo que contribuye la otra. Por lo tanto, al considerar correlaciones comunes entre cada una de las variables de predicción y la variable dependiente, la importancia relativa de cada una de las variables puede ser muy diferente. En este ejemplo, se observa que las correlaciones comunes muestran un patrón similar, aunque la diferencia entre las dos variables no es tan importante como cuando se consideran los coeficientes beta.

Otra información importante en la tabla es R^2. Este número indica la proporción de error cuadrático de las predicciones, que se reduce al utilizar esta regla de predicción óptima, con respecto a utilizar sólo el promedio de aceptación por parte de sus pares para predecir cada registro. Es una forma estándar de describir la calidad de la regla de predicción óptima. En este caso, la reducción proporcional de error es del 14%. Además, la reducción proporcional de error es exactamente el cuadrado de la correlación total. La raíz cuadrada de 0,14 es 0,37. Por lo tanto, la correlación total (denominada **correlación múltiple** y simbolizada con la R mayúscula) de la aceptación por parte de los pares con respecto a la enseñanza no social y al entrenamiento social tomados en su conjunto es $R = 0{,}37$.

7. a) $\hat{Z}_{\text{Aceptación por parte de los pares}}$
$= (0{,}10)(Z_{\text{Enseñanza no social}}) + (0{,}32)(Z_{\text{Entrenamiento social}})$

A $(0{,}10)(-2) + (0{,}32)(0) = -0{,}20$
B $(0{,}10)(0) + (0{,}32)(0) = 0$
C $(0{,}10)(2) + (0{,}32)(0) = -0{,}20$
D $(0{,}10)(0) + (0{,}32)(-2) = -0{,}64$
E $(0{,}10)(0) + (0{,}32)(2) = -0{,}64$
F $(0{,}10)(-2) + (0{,}32)(2) = -0{,}20 + 0{,}64 = -0{,}84$
G $(0{,}10)(-1) + (0{,}32)(-2) = -0{,}10 - 0{,}64 = -0{,}74$

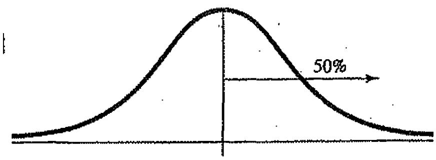
b) $\hat{Z}_{\text{Aceptación por parte de los pares}}$
$= (0{,}27)(Z_{\text{Estilo de reacción}}) + (0{,}29)(Z_{\text{Entrenamiento Social}})$

A $(0{,}27)(-2) + (0{,}29)(0) = -0{,}54$
B $(0{,}27)(0) + (0{,}29)(0) = 0$
C $(0{,}27)(2) + (0{,}29)(0) = -0{,}54$
D $(0{,}27)(0) + (0{,}29)(-2) = -0{,}58$
E $(0{,}27)(0) + (0{,}29)(2) = 0{,}58$
F $(0{,}27)(2) + (0{,}29)(2) = 0{,}54 + 0{,}58 = 1{,}12$
G $(0{,}27)(-1)+(0{,}29)(-2) = -0{,}27 - 0{,}58 = -0{,}85$

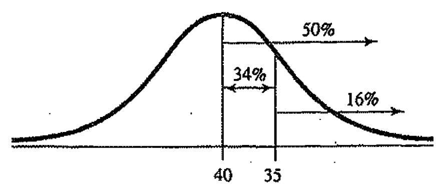
Capítulo 5

1. a) 50%, b) 16%, c) 98%, d) 84%, e) 50%, f) 84%, g) 2%, h) 16%; i) 50, j) 45, k) 40, l) 35, m) 30.

Nota: Será mucho más fácil resolver problemas como éste trazando un dibujo de la curva normal y marcándola como muestran los dibujos a continuación.



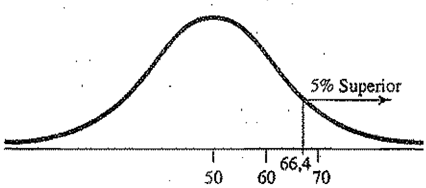
(a)



(b)

2. a) De la tabla de áreas bajo la curva normal en el apéndice B, el 43% (0,43) presentan puntuaciones Z entre la media y 1,5. Por definición, el 50% presentan puntuaciones Z por debajo de la media. Por lo tanto, el porcentaje total por debajo de 1,5 es 50% + 43% = 93%.
b) El 43% presentan puntuaciones Z entre 1,5 y la media, y dado que hay un 50% total por encima de la media, 50% − 43% = 7% quedan por encima de 1,5 (o puede restar el 93% obtenido en el ejercicio 2a y calcular que sólo queda el 7% del total de 100%).
c) El 43% de los valores se encuentran entre la media y 1,5, y la curva normal es simétrica. Por lo tanto, un 43% están entre la media y −1,5. Dado que un 50% se encuentra debajo de la media, esto deja un 7% por debajo de −1,5.
d) −93%, e) 2%, f) 98%, g) 33%, h) 4%, i) 5%.
(Una vez más, todos estos problemas son más fáciles de resolver trazando un dibujo de la curva normal y marcando las áreas).

3. a) El 10% superior significa que un 90% está por debajo de él. De este 90%, el 50% está por debajo de la media. Por lo tanto, el 10% superior es aquel punto entre el cual se ubica el 40% de los valores y la media. Buscando 40,0 en la tabla de áreas de la curva normal (el número más cercano es 39,97), descubrimos que éste equivale a una puntuación Z de +1,28.
   b) 2,33.
4. Puntuación Z necesaria = 1,64, que corresponde a una puntuación original de 50 + (10)(1,64)= 66,4. Explicación: los valores correspondientes a casi todo lo que midamos en la naturaleza y en la psicología tienden a seguir aproximadamente el patrón particular que mostramos debajo, denominado modelo normal. En un modelo normal, la mayoría de los valores están cerca del medio con cantidades menores pero iguales de valores a cada extremo. Dado que la curva normal está matemáticamente definida, se puede calcular la proporción exacta de valores en cualquier intervalo determinado, y esas proporciones han sido calculadas y presentadas en tablas especiales.



Las tablas de áreas bajo la curva normal se basan en lo que se denominan puntuaciones Z. Las puntuaciones Z, a su vez, se basan en la media y el desvío estándar. La media es el promedio común, la suma de todos los valores dividida por la cantidad de ellos. El desvío estándar es una medida de la dispersión de un grupo de valores. Sin detenernos en detalles, indica la cantidad promedio en que los valores difieren del promedio. (Para ser exactos, es la raíz cuadrada del promedio de los cuadrados de las diferencias de cada valor con el promedio). La puntuación Z es la cantidad de desvíos estándar que separan a un valor de la media. La tabla de áreas bajo la curva normal indica el porcentaje de casos que están incluidos entre la media y cualquier puntuación Z en particular.

Dado que sabemos que los valores de la prueba de coordinación siguen una distribución normal, podemos buscar en la tabla la puntuación Z, que corresponde al punto entre el cual se encuentra el 45% de los registros, la media (debido a que la curva normal es completamente simétrica, el 50% de los valores se encuentran por sobre la media, quedando un 5% superior al 45%). Esto resulta ser una puntuación Z de 1,64 (en realidad, no existe un punto exacto en la tabla para el 45%, por lo cual podríamos haber tomado tanto el 1,64 como el 1,65).

Si el desvío estándar es igual a 10, una puntuación Z de 1,64 está a 16,4 puntos por encima de la media. Sumando ese resultado a la media 50 obtenemos un valor mínimo necesario de 66,4 para estar en el 5% superior.

5. a) 10/50: p = 10/50 = 0,2; b) .4; (c) (10 + 20)/50 = 0,6; d) 0,6; e) 1.
6. Una muestra es un grupo de personas analizadas que representan el grupo completo al que se pretende aplicar los resultados obtenidos, y que se denomina población. (En este caso, la población está formada por todos los alumnos secundarios australianos). Se estudia una muestra porque sería muy poco práctico o imposible estudiar la población completa.

   Una forma de asegurarse que una muestra no es sistemáticamente no representativa, es seleccionar la muestra al azar. Esto no significa hacerlo de forma casual. Por ejemplo, tomar sólo a los alumnos que están más disponibles para participar en la prueba sería un muestreo casual. Pero este no sería un buen método porque cualesquiera que sean los factores que hicieron que esas personas estuvieran disponibles, como por ejemplo vivir en una ciudad cercana, podrían hacer que no fueran representativos de la población en su totalidad. Un ejemplo de una verdadera selección al azar sería conseguir una lista de todos los alumnos secundarios de Australia, numerar a cada alumno y luego utilizar una tabla de números aleatorios para escoger la cantidad que se necesiten para realizar el estudio.

Capítulo 6

1. a) Una hipótesis de investigación es el enunciado de la relación predicha entre poblaciones (por ejemplo, que tendrán diferentes medias).
   b) La hipótesis nula es el enunciado de una relación entre poblaciones opuestas, a lo que se predice a través de la hipótesis de investigación (por ejemplo, que las dos poblaciones tienen la misma media).
   c) La prueba de hipótesis es el procedimiento estadístico lógico para analizar la verosimilitud de un patrón de resultados particular logrado a través de un determinado estudio bajo el supuesto de que la hipótesis nula es verdadera.

d) La distribución comparativa es la distribución correspondiente a la situación en la que la hipótesis nula es verdadera; es la distribución con la cual comparamos el valor de la muestra.

e) El "nivel de significación de 0,05" es aquella situación que se da en la prueba de hipótesis, en la que decidimos rechazar la hipótesis nula porque la probabilidad de obtener esos resultados en particular, si la hipótesis nula fuera verdadera, es menor al 5%.

f) Una prueba de una cola es un procedimiento de prueba de hipótesis en el que la hipótesis de investigación especifica una dirección particular de la diferencia (por ejemplo, que la media de una población será mayor que la de la otra).

2. i) a)
Población 1: niños canadienses hijos de bibliotecarios.
Población 2: todos los niños canadienses.
b) Hipótesis de investigación: los niños de la población 1 presentan un promedio más alto de habilidad para la lectura que los niños de la población 2.
c) Hipótesis nula: el promedio de habilidad para la lectura de la población 1 no es mayor que el de la población 2.
d) Una cola, porque la cuestión es si lo "hacen mejor"; por lo tanto, nos interesa la diferencia sólo en una dirección.

ii) a)
Población 1: personas que viven en determinada ciudad.
Población 2: todas las personas que viven en la región.
b) Hipótesis de investigación: las poblaciones 1 y 2 presentan diferentes medias de ingresos.
c) Hipótesis nula: las poblaciones 1 y 2 presentan la misma media de ingresos.
d) Dos colas, porque la cuestión es si los ingresos de las personas de la ciudad son "diferentes" a los ingresos de los que viven en toda la región; por lo tanto, nos interesa una diferencia en cualquier dirección.

iii) a) Población 1: personas que han sufrido la experiencia de un terremoto.
Población 2: personas en general.
b) Hipótesis de investigación: las poblaciones 1 y 2 presentan diferentes medias en cuanto a niveles de confianza en sí mismos.
c) Hipótesis nula: las poblaciones 1 y 2 presentan la misma media en cuanto al nivel de confianza en sí mismos.

d) Dos colas, porque podrían tener más como menos confianza en sí mismos.

3.

Estudio	Corte	Puntuación Z en la distribución	Decisión
A	+1,64	2,0	Rechaza la hipótesis nula
B	±1,96	2,0	Rechaza la hipótesis nula
C	+2,33	2,0	No concluyente
D	±2,57	2,0	No concluyente
E	+1,64	1,0	No concluyente
F	±2,57	4,0	Rechaza la hipótesis nula
G	±2,57	3,0	Rechaza la hipótesis nula
H	±2,57	2,0	No concluyente
I	−1,64	−2,0	Rechaza la hipótesis nula

4. Los cinco pasos de la prueba de hipótesis son:

1. Replantear el problema en función de hipótesis de investigación e hipótesis nula de las poblaciones. Las dos poblaciones de interés son:

Población 1: alumnos que han evitado utilizar su sentido del olfato.
Población 2: alumnos en general.

La hipótesis de investigación supone que los alumnos que evitaron utilizar su sentido del olfato (población 1) tendrán un rendimiento más bajo que los alumnos en general (población 2) en la prueba del sentido del gusto. La hipótesis nula supone que los alumnos que evitaron utilizar sus sentidos del olfato (población 1) no tendrán un rendimiento más bajo en la prueba que los alumnos en general (población 2).

2. Determinar las características de la distribución comparativa. La distribución comparativa será igual a la población 2. Según se establece en el enunciado del problema, $\mu = 14$ y $\sigma = 4$. Supondremos que presenta una curva normal.

3. Determinar el punto muestral de corte en la distribución comparativa, en el que debería rechazarse la hipótesis nula. Con un nivel de 0,05 para una prueba de una cola, el corte se ubica en −1,64.

4. Determinar el valor muestral en la distribución comparativa. El valor muestral era 5.

$$Z = (5-14)/4 = -9/4 = -2,25.$$

5. Comparar los valores de los pasos 3 y 4 para decidir si se rechaza o no la hipótesis nula. Una puntuación Z de −2,25 se encuentra por debajo del punto de corte establecido en −1,64. Por lo tanto, podemos rechazar la hipótesis nula. Se sostiene la hipótesis de investigación: no utilizar el

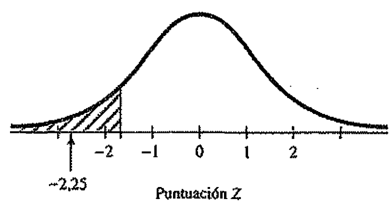
sentido del olfato provoca menos identificaciones correctas.

Explicación: resumiendo, este problema se resuelve considerando la posibilidad de que ocurra el escenario en el que no tener el sentido del olfato no produce diferencia alguna. Si el sentido del olfato no produjera ninguna diferencia, la probabilidad de que el alumno analizado obtenga una cantidad determinada de identificaciones correctas es simplemente igual a la probabilidad de que los alumnos en general obtengan una cantidad determinada de identificaciones correctas. Y dado que conocemos la distribución de la cantidad de identificaciones correctas de los alumnos en general, esa probabilidad puede ser determinada, y sucede que sería bastante improbable obtener sólo 5 identificaciones correctas; por lo tanto, el investigador concluye que la falta del sentido del olfato produce una diferencia.

Para entrar un poco más en detalles, el punto clave es determinar las probabilidades. Suponemos que la cantidad de identificaciones correctas de los alumnos en general se distribuye en forma normal: se trata de un patrón matemático específico, la curva normal, a veces denominada "forma de campana", en la que en la mayoría de los casos se ubican en el medio y van disminuyendo progresivamente a medida que los números aumentan o disminuyen. Existen tablas que muestran exactamente qué proporción de casos se ubica entre el medio y cualquier punto en particular de la curva normal. Estas tablas utilizan "puntuaciones *Z*", versiones transformadas de los valores originales, que representan la cantidad de desvíos estándar por encima de la media. La media es el promedio ordinario (la suma de los valores dividida por la cantidad de valores). El desvío estándar puede considerarse como la cantidad promedio en la que los valores difieren de la media. (Estrictamente hablando, es la raíz cuadrada del promedio de los cuadrados de la diferencia de cada valor con respecto a la media).

Al evaluar el resultado de un experimento, muchos investigadores utilizan la norma convencional que establece que si un resultado podría haber ocurrido menos de un 5% de las veces bajo un escenario determinado, ese escenario será considerado improbable. Las tablas de áreas bajo la curva normal muestran que el 5% superior de la curva normal comienza con una puntuación *Z* de 1,64. Dado que la curva normal es completamente simétrica, el 5% inferior incluye a todos las puntuaciones *Z* debajo de −1,64. Por lo tanto, incluso antes de realizar el experimento, el investigador probablemente establezca la siguiente regla: el escenario en el que la falta del sentido del olfato no produce diferencias será rechazado como improbable si la cantidad de identificaciones correctas (convertidas a puntuanción *Z* utilizando la media y el desvío estándar correspondiente a los alumnos en general) es menor a −1,64.

La cantidad real de identificaciones correctas del alumno que no pudo utilizar el sentido del olfato fue 5. Se nos dice que la curva normal correspondiente a los alumnos en general presentaba una media de 14 y un desvío estándar de 4. Cinco identificaciones correctas implican 9 por debajo de la media de 14; en términos de desvíos estándar de 4 por unidades cada uno, implica 9/4 desvíos por debajo de la media, es decir, una puntuación *Z* de −2,25.
Dado que −2,25 es menor que −1,64, el investigador concluye que el escenario en el que la falta del sentido del olfato no produce efectos es improbable. El gráfico que sigue ilustra el problema:



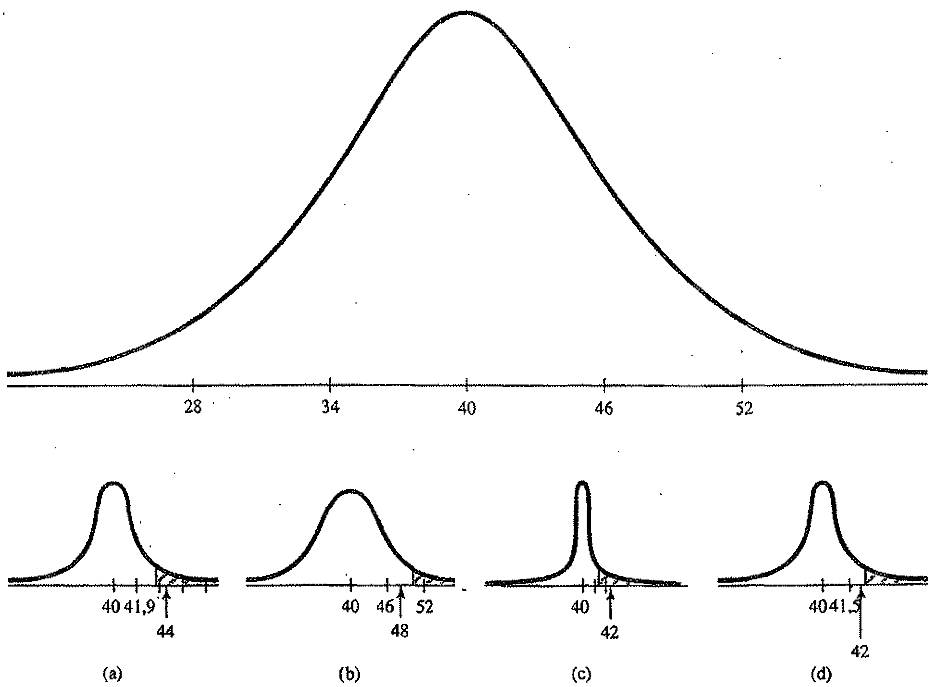
5. El punto de corte (con nivel 0,01 para una prueba de una cola) es igual a −2,326; la puntuación *Z* en la distribución comparativa correspondiente al paciente estudiado es +1,2; el experimento no es concluyente.

   Los cinco pasos de la prueba de hipótesis y la explicación son similares a los del ejercicio 4, excepto que la explicación puede ser más breve ya que el resultado es contrario a la hipótesis, y no es necesario explicar *M*, *SD* y *Z*.
6. Las dos "*Ms*" (5,7 y 4,8) y la "$p < 0,05$" son cruciales. *M* significa **media**, el promedio de los valores de un determinado grupo. La cantidad promedio de veces por día que los participantes con alto grado de narcisismo se miraron al espejo fue de 5,7, mientras que el promedio para los participantes con bajo grado de narcisismo fue de sólo 4,8. La "$p<0,05$" indica que esta diferencia es estadísticamente significativa al nivel 0,05, es decir, si el nivel de narcisismo de una persona no produjera ninguna diferencia en cuanto a la frecuencia con que esa persona se mira en el espejo, las posibilidades de encontrar dos grupos de participantes que presentaran esta diferencia en cuanto a mirarse en

el espejo sólo por casualidad, sería menor a 0,05 (menos del 5%). Por lo tanto, rechazamos esa posibilidad por improbable y concluimos que el nivel de narcisismo efectivamente produce una diferencia en cuanto a la frecuencia con que una persona se mira al espejo.

Capítulo 7

1. El desvío estándar de la distribución de medias es menor al desvío estándar de la distribución poblacional de individuos, porque existe menos variación entre medias de muestras formadas por más de un valor que entre valores individuales. Existe menos variación porque la probabilidad de que dos registros extremos en la misma dirección sean elegidos al azar para formar la misma muestra es menor que la probabilidad de que cada uno de estos valores extremos sea elegido individualmente.
2. a) $\sigma^2 = 10^2 = 100$; $\sigma^2_M = \sigma^2/N = 100/2 = 50$; $\sigma_M = \sqrt{\sigma^2_M} = \sqrt{50} = 7{,}07$ b) 5,77; c) 5; d) 4,47; e) 3,16; f) 2,24; g) 1.
3. a) Límite superior $= M + (\sigma_M)(1{,}96) = 100 + (7{,}07)(1{,}96) = 113{,}86$; límite inferior $= 100 + (7{,}07)(-1{,}96) = 86{,}14$. b) 111,31, 88,69; c) 109,8, 90,2; d) 108,76, 91,24; e) 106,19, 93,81; f) 104,39, 95,61; g) 101,96, 98,04.
4. Dado que la distribución de la población de individuos es normal, también lo será la distribución de medias. Por lo tanto, basándonos en la tabla de áreas bajo la curva normal, se necesita una puntuación Z de al menos 1,64 para estar dentro del 5% superior. Para la muestra, a: $\sigma_M = \sqrt{(36/10)} = 1{,}90$. Z (en la distribución de medias) $= (44 - 40)/1{,}90 = 4/1{,}90 = 2{,}11$. Dado que 2,11 es más extremo que 1,64, la muestra tiene menos de un 5% de probabilidades. La respuesta final para el punto b tiene más de un 5% de probabilidades; para el punto c, menos probabilidades y



para el punto d, más del 5% de probabilidades. Las distribuciones están representadas en el gráfico que sigue a continuación.

5. a) Los cinco pasos de la prueba de hipótesis:

1. Replantear el problema en función de hipótesis de investigación e hipótesis nula sobre poblaciones. Las dos poblaciones son:

Población 1: mujeres mayores que reciben el programa especial.

Población 2: mujeres mayores en general (que no reciben el programa especial).

La hipótesis de investigación establece que la población de mujeres mayores que reciben el programa especial (población 1) tendrán un tiempo de reacción más rápido que las mujeres mayores en general (población 2). La hipótesis nula establece que los valores de la población 1 no serán menores que los de la población 2.

2. Determinar las características de la distribución comparativa. La distribución comparativa es una distribución de medias de muestras formadas por 25 valores, tomados de la distribución de la población 2. $\mu = 1{,}8$; $\sigma^2_M = \sigma^2/N = 0{,}5^2/25 = 0{,}25/25 = 0{,}01$; $\sigma_M = \sqrt{0{,}01} = 0{,}1$. Dado que la población es normal, la distribución de medias es normal.

3. Determinar los valores muestrales de corte en la distribución comparativa a partir de los cuales debería rechazarse la hipótesis nula. Utilizando una prueba de una cola (los investigadores predijeron un menor tiempo de reacción) a nivel 0,01, el punto de corte es –2,33.

4. Determinar el valor muestral en la distribución comparativa $Z = (1{,}5-1{,}8)/0{,}1 = -0{,}3/0{,}1 = -3$.

5. Comparar los registros de los pasos 3 y 4 para decidir si se rechaza o no la hipótesis nula. Dado que –3 es menor a –2,33, podemos rechazar la hipótesis nula. Se sostiene la hipótesis de investigación; las mujeres mayores que participaron del programa especial muestran un menor tiempo promedio de reacción.

b) Intervalo del 99% de certeza: límite superior $= M + (\sigma_M)(2{,}57) = 1{,}5 + (0{,}1)(2{,}57) = 1{,}5 + 0{,}257 = 1{,}76$; límite inferior $= 1{,}5 + (0{,}1)(-2{,}57) = 1{,}24$.

c) Explicación: se trata de un problema estándar de prueba de hipótesis, excepto por el hecho de que no podemos comparar directamente el tiempo de reacción del grupo de las 25 mujeres analizadas con la distribución de los tiempos de reacción de mujeres individuales en general. Esto ocurre porque la distribución de mujeres en general es una distribución de valores individuales, y nosotros tenemos el promedio de un grupo de puntuaciones de 25 personas. La probabilidad de que un grupo de valores tenga una media extrema, sólo por azar, es mucho menor a la probabilidad de que cualquier individuo tenga un valor extremo sólo por azar. (Esto ocurre porque al seleccionar valores al azar cuando seleccionamos varios valores, cualquier valor extremo probablemente será equilibrado por valores menos extremos o extremos en dirección opuesta). Por lo tanto, la distribución adecuada para comparar la media de los tiempos de reacción del grupo de 25 personas es la distribución que resultaría si seleccionamos al azar muchas series de 25 valores de tiempos de reacción y calculáramos la media de cada serie de 25 valores.

Tal distribución, formada por diversas medias muestrales, tiene la misma media que la distribución original de valores individuales (no existe razón para que sea de otro modo), pero es una curva más estrecha ya que las posibilidades de que existan extremos es menor. De hecho, se sabe matemáticamente que su varianza será exactamente la varianza de la distribución original de observaciones individuales dividida por la cantidad de valores de cada muestra. En este caso, tendremos una distribución de medias con una media de 1,8 y un desvío estándar de 0,1 ($\sqrt{(5^2/25)}$), y será, además, una distribución normal porque una distribución de diversas medias tomadas de una población normalmente distribuida también es normal.

El punto de corte correspondiente al nivel de significación 0,01, y una prueba de una cola, es –2,33. La clasificación media del grupo de 25 mujeres que recibieron el programa especial, 1,5, estaba a 3 desvíos estándar por debajo de la media de la distribución de medias, siendo claramente más extrema que el punto de corte. Por lo tanto, podemos rechazar la hipótesis nula y concluir que la información sostiene la hipótesis que establece que las mujeres mayores que participan del programa especial demuestran tener menores tiempos de reacción. El intervalo de confianza es una estimación del conjunto de valores que probablemente incluya la verdadera media poblacional del grupo estudiado (población 1: en este problema, las mujeres que reciben el programa especial para mejorar el tiempo de reacción). Un intervalo del 99% de confianza es el conjunto de valores que nos da un 99% de certeza de incluir la verdadera media poblacional. Para calcular los límites superior e inferior del intervalo, comenzamos por considerar que la mejor estimación puntual de la media de la población 2 es la media de la muestra tomada de esa población (en este caso, 1,5). También suponemos que el desvío estándar de la distribución de medias para esta población 2 es el mismo que el de la población conocida (que ya calculamos en 0,1). Basándonos en esta información, si la media real de la población era 1,5, el 99% de las veces las medias muestrales se ubicarían entre una puntuación Z de +2,57 (el punto en la curva normal que incluye 49,5% de los registros por sobre la media) y –2,57. En nuestro ejemplo, estas puntuaciones Z corresponden a las puntuaciones originales 1,76 y 1,24.

Los valores que hemos calculado son los límites del intervalo de confianza. ¿Por qué? Supongamos que la media real de la población fuera 1,76. En ese caso, existiría un 0,5% de posibilidades de obtener una media tan baja o más baja que 1,5. (Es decir, con una media de 1,76 y un desvío estándar de 0,1, 1,5 está exactamente 2,57 desvíos estándar por debajo de la media, que es el punto que corresponde al corte del 0,5% inferior). De modo similar, si la media real de la población fuera 1,24, habría sólo un 0,5% de posibilidades de obtener una media mayor a 1,5.

6. a) Los cinco pasos de la prueba de hipótesis deberían realizarse de manera similar a los descriptos en el problema 5 anterior. La información clave para este problema es que la distribución de medias estará normalmente distribuida con una media de 5,5 y un desvío estándar de 0,2 (es decir, $\sqrt{8^2/16}$). Utilizando una prueba de una cola a un nivel de 0,05, el punto de corte necesario es 1,64. La puntuación media muestral de 5,9 se ubica, en esta distribución, 2 desvíos estándar por encima de la media: (5,9 – 5,5)/0,2. Por lo tanto, se rechaza la hipótesis nula.

b) Intervalo del 95% de confianza: límite superior = $M + (\sigma_M)(1,96) = 5,9 + (0,2)(1,96) = 5,9 + 0,392 = 6,29$; límite inferior $= 5,9 + (0,2)(-1,96) = 5,51$.

c) La descripción para una persona que nunca ha estudiado estadística sería similar a la descripción que aparece en la respuesta al problema 4 del capítulo 6, más el material adicional de la respuesta al problema 5 de este capítulo.

Capítulo 8

1. Alfa es la probabilidad de rechazar la hipótesis nula cuando es verdadera. Beta es la probabilidad de

2.

Conclusión arrojada por la prueba de hipótesis	Condición real de la hipótesis de investigación: Verdadera	Condición real de la hipótesis de investigación: Falsa
(a) Rechazar la nula	Decidir que un mayor descanso mejora el comportamiento es la decisión correcta; de hecho, lo hace.	Decidir que mayor descanso mejora el comportamiento, decisión incorrecta; de hecho, no lo hace.
No concluyente	Decidir que se desconoce la relación entre descanso y comportamiento, decisión inadecuada; de hecho, un mayor descanso mejora el comportamiento.	Decidir que se desconoce la relación entre descanso y comportamiento, decisión adecuada; de hecho, no están relacionados.
(b) Rechazar la nula	Decidir que los daltónicos distinguen mejor, decisión correcta; de hecho lo hacen.	Decidir que los daltónicos distinguen mejor, decisión incorrecta; de hecho no lo hacen.
No concluyente	Decidir que se desconoce si los daltónicos distinguen mejor, decisión inadecuada; de hecho lo hacen.	Decidir que se desconoce si los daltónicos distinguen mejor, decisión adecuada; de hecho no lo hacen.
(c) Rechazar la nula	Decidir que los individuos que han asistido a psicoterapia son más tolerantes, decisión correcta; lo son.	Decidir que individuos que han asistido a psicoterapia son más tolerantes, decisión incorrecta; no lo son.
No concluyente	Decidir que se desconoce si son o no más tolerantes, decisión inadecuada; lo son.	Decidir que se desconoce si son o no más tolerantes, decisión adecuada; no lo son.

no rechazar la hipótesis nula cuando en realidad la hipótesis nula es falsa.

3. *Véase* la tabla en la parte superior de la siguiente columna y la figura en la parte superior de la próxima página.

4. Z necesario para significación = 1,64; $\sigma_M = 2$ (es decir, $\sqrt{[144/36]} = 2$); puntuación original necesaria para significación = 53,28 (es decir, 50 + [1,64][2] = 53,28); puntuación Z correspondiente en la distribución predicha = –0,86 (es decir, [53,28 – 55]/2 = –0,86); según la tabla Z, beta = 0,19; potencia = 0,81 (es decir, 1 – 0,19).

 Explicación: La potencia es la posibilidad de rechazar la hipótesis nula si la hipótesis de investigación es verdadera. Para calcular la potencia, el primer paso es determinar las características de la distribución comparativa. En este experimento, será una distribución de medias (de muestras de 36 artistas cada una) que está normalmente distribuida (ya que la población lo está) con una media de 50 y un desvío estándar de 2 (según los cálculos descriptos anteriormente). Para rechazar la hipótesis nula, la puntuación Z de la media muestral debe ser superior a 1,64 (se trata de una prueba de una cola a nivel 0,05), lo que corresponde a una media muestral de 53,28 puntos originales.

 Ahora bien, los cálculos de la potencia son los siguientes. El investigador elaboró la hipótesis de que la media poblacional de artistas es 55 (e, implícitamente, que esta población también es normal con la misma σ de 12). La distribución de medias de esta población sería normal con una media = 55 y $\sigma_M = 2$. Ya determinamos que cualquier media por sobre 53,28 será significativa en términos de la distribución comparativa. Pero un valor de 53,28 se corresponde con una puntuación Z de sólo –0,86 en la distribución basada en la hipótesis del investigador. Utilizando la tabla de áreas bajo la curva normal, el 81% del área bajo la curva se encuentra por encima de este punto. Suponiendo que las predicciones del investigador sean correctas, existe un 81% de posibilidades de que una muestra de 36 artistas produzca un resultado lo suficientemente alto como para rechazar la hipótesis nula. Es decir, la potencia es del 81%.

 Las dos distribuciones de medias involucradas y las áreas de significación y potencia están representadas gráficamente al final de la siguiente página.

	Z necesario para obtener significación	σ_M	Valor para obtener significación
(a)	1,64	0,4	90,66
(b)	1,64	0,4	90,66
(c)	1,64	0,2	90,33
(d)	1,64	1,0	91,64
(e)	2,33	0,4	90,93
(f)	1,96	0,4	90,78

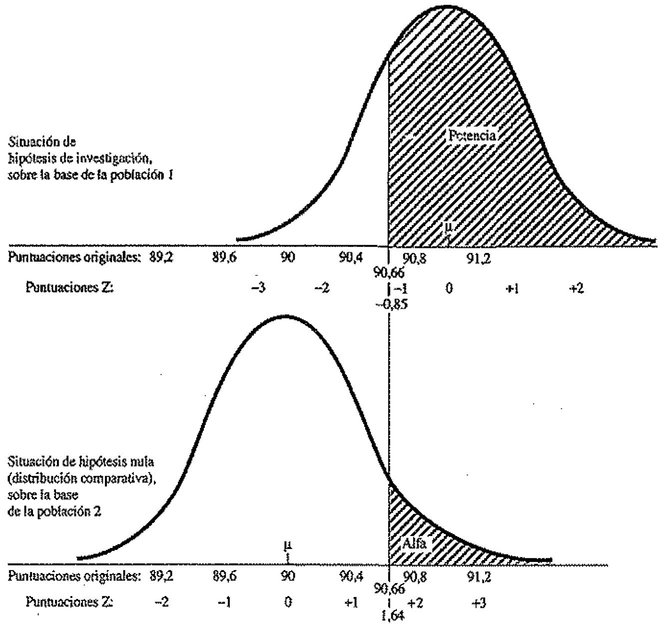
Z para significación en la población predicha	Beta	Potencia	Tamaño del efecto
(a)(90,66 – 91)/0,4 = –0,85	0,20	0,80	1/4
(b)(90,66 – 92)/0,4 = –3,35	<0,01	>0,99	1/2
(c)(90,33 – 91)/0,2 = –3,35	<0,01	>0,99	1/2
(d)(91,64 – 91)/ 1 = 0,64	0,74	0,26	1/4
(e)(90,93 – 91)/0,4 = –0,18	0,43	0,57	1/4
(f) (90,78 – 91)/0,4 = –0,55	0,29	0,71	1/4

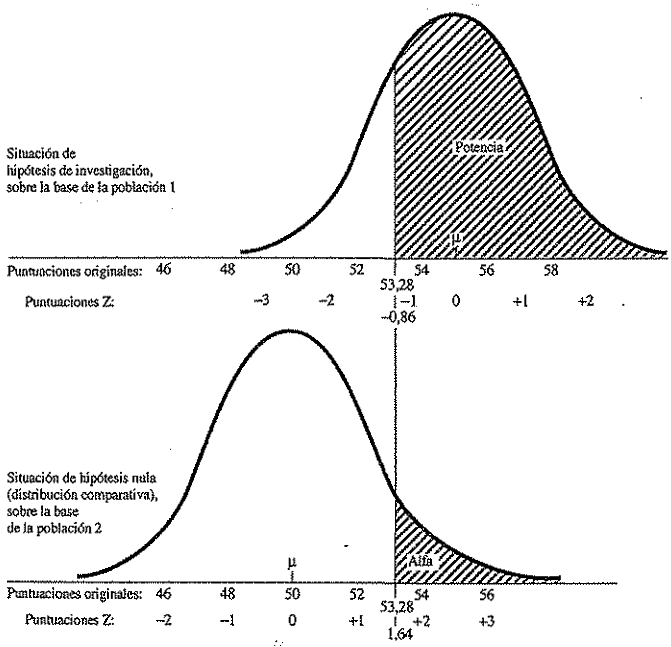
El dibujo de las distribuciones superpuestas para la versión a aparece en la parte superior de la página 634.

5. a) No la afecta (eso es lo que prueba el nivel de significación).
 b) Probablemente de poca importancia (debido al pequeño tamaño del efecto).

6. a) Aumenta la potencia; b) disminuye la potencia; c) aumenta la potencia; d) disminuye la potencia; e) disminuye la potencia.

7. a) Al planificar un experimento, para permitir cambios de distinto tipo (o incluso abandonar el proyecto) si la potencia es demasiado baja. (O posiblemente hacer que el proyecto sea menos costoso, por ejemplo, reduciendo la cantidad de participantes si la potencia es más alta de lo razonablemente necesario).
 b) Después de realizar un estudio que ha arrojado resultados no concluyentes, para evaluar si la falla del estudio debería atribuirse a que la hipótesis nula es falsa (en el caso de que la potencia sea alta) o a una potencia inadecuada, de forma tal que aún es razonable pensar que futuras investigaciones podrían tener la posibilidad de ser significativas. (Además, en el caso de un resultado significativo con un gran tamaño de muestra, si la potencia es muy alta, esto sugiere que es posible un bajo tamaño del efecto indicando que, aunque el resultado es significativo, puede no ser muy importante).


Situación de hipótesis de investigación, sobre la base de la población 1
Potencia
μ
Puntuaciones originales: 89,2 89,6 90 90,4 90,8 91,2
90,66
Puntuaciones Z: −3 −2 −1 0 +1 +2
−0,85
Situación de hipótesis nula (distribución comparativa), sobre la base de la población 2
μ
Alfa
Puntuaciones originales: 89,2 89,6 90 90,4 90,8 91,2
90,66
Puntuaciones Z: −2 −1 0 +1 +2 +3
1,64


Situación de hipótesis de investigación, sobre la base de la población 1
Potencia
μ
Puntuaciones originales: 46 48 50 52 54 56 58
53,28
Puntuaciones Z: −3 −2 −1 0 +1 +2
−0,86
Situación de hipótesis nula (distribución comparativa), sobre la base de la población 2
μ
Alfa
Puntuaciones originales: 46 48 50 52 54 56
53,28
Puntuaciones Z: −2 −1 0 +1 +2 +3
1,64

Capítulo 9

1. a) t necesario ($gl = 63, p < 0{,}05$, una cola) $= -1{,}671$.
$S_M = (\sqrt{S^2/N}) = (\sqrt{9/64}) = \sqrt{0{,}141} = 0{,}38$.
$t = (M - \mu)/S_M = (11 - 12{,}40)/0{,}38$
$= -1{,}40/0{,}38 = -3{,}68$.
Se rechaza la hipótesis nula.

   b) t necesario $= 2{,}690$; $S_M = 2{,}55$; $t = 1{,}32$; no se rechaza la hipótesis nula.

   c) t necesario $= 2{,}364$; $S_M = 0{,}13$; $t = 3{,}15$; se rechaza la hipótesis nula.

2. a) Pasos de la prueba de hipótesis:

   1. Replantear el problema en función de hipótesis de investigación e hipótesis nula de poblaciones.

   Población 1: tiempo de respuesta con el nuevo jefe de policía.

   Población 2: tiempo de respuesta con el antiguo jefe de policía.

   La hipótesis nula establece que las dos poblaciones son iguales. La hipótesis de investigación establece que las dos poblaciones son diferentes.

   2. Determinar las características de la distribución comparativa..

   Población 2: forma = se presume normal; $\mu = 30$; σ^2 = desconocida;
   $S^2 = \Sigma(X - M)^2/(N - 1) = SS/gl$
   $= 124/(10 - 1) = 13{,}78$.

   Distribución de medias: forma $= t$ ($gl = 9$);
   $\mu_M = 30$;
   $S_M = \sqrt{(S^2/N)} = \sqrt{(13{,}78/10)} = 1{,}17$

   3. Determinar el punto muestral de corte en la distribución comparativa, a partir del cual debería rechazarse la hipótesis nula.

   t necesario ($gl = 9, p < 0{,}05$, una cola)
   $= -1{,}833$.

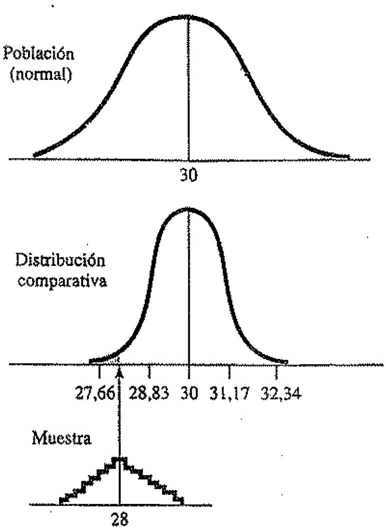
   4. Determinar el valor muestral en la distribución comparativa.

   La media del nuevo jefe de policía es
   $M = \Sigma X/N = 280/10 = 28$.
   $t = (M - \mu)/S_M = (28 - 30)/1{,}17 = -1{,}71$

   5. Comparar los registros de los pasos 3 y 4 para decidir si se rechaza o no la hipótesis nula.

   El registro en 4 ($-1{,}71$) no es más extremo que el registro en 3 ($-1{,}833$). Por lo tanto, no se rechaza la hipótesis nula.

   b)



   c) Explicación: es la misma que la del problema 4, serie 1, de este capítulo, excepto que en lugar de diferencias, aquí se utilizan los valores reales, y la media poblacional esperada son los 30 minutos (1/2 hora) que el jefe de policía había prometido cuando era candidato.

3. a) t necesario ($gl = 19$, $p < 0{,}05$, una cola) $= 1{,}729$.
$S_M = \sqrt{(S^2/N)} = \sqrt{(8.29/20)} = \sqrt{415} = 0{,}64$.
$t = (M - \mu)/S_M = (1{,}7 - 0)/0{,}64 = 2{,}66$.
Se rechaza la hipótesis nula.
$d = M/S = 1{,}7/\sqrt{8{,}29} = 0{,}59$.

   b) t necesario $= \pm 1{,}980$; $S_M = \sqrt{141{,}53/164} = 1{,}59$; $t = (2{,}3 - 0)/1{,}59 = 1{,}45$; no se rechaza la hipótesis nula; $d = 0{,}11$.

   c) t necesario $= -2{,}624$; $S_M = 0{,}52$; $t = -4{,}23$; se rechaza la hipótesis nula; $d = -1{,}1$.

4. a) Pasos de la prueba de hipótesis:

   1. Replantear el problema en función de hipótesis de investigación e hipótesis nula de poblaciones.

   Población 1: ciudades como aquellas que participaron en el programa para reducir los residuos.

   Población 2: ciudades que no cambian en cuanto a cantidad de residuos durante un periodo de un año.

   La hipótesis de investigación establece que la población 1 presenta una media de disminución de residuos superior a la de la población 2. La hipótesis nula establece que la población 1 no tiene una media de disminución de residuos mayor que la de la población 2.

2. Determinar las características de la distribución comparativa.

Forma de la población 2 = se presume normal; $\mu = 0$; σ^2 = desconocido; $S^2 = 50/3 = 16{,}67$; Forma de la distribución de medias = t ($gl = 3$); $\mu_M = 0$; $S_M = \sqrt{S2/N} = \sqrt{16{,}67/4} = \sqrt{4{,}17} = 2{,}04$

3. Determinar los valores muestrales de corte en la distribución comparativa, a partir de los cuales debería rechazarse la hipótesis nula. t necesario ($gl = 3, p < 0{,}01$, una cola) $= 4{,}541$.

4. Determinar el valor muestral en la distribución comparativa.

Valores de cambio = 7, 6, –1, 8;

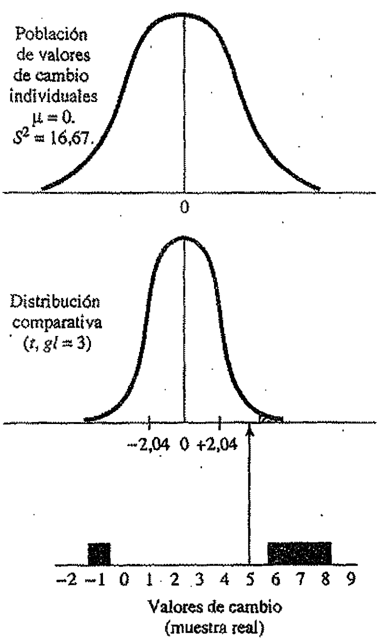
$M = 20/4 = 5$; $t = (5 - 0)/2{,}04 = 2{,}45$

5. Comparar los valores de los pasos 3 y 4 para decidir si se rechaza o no la hipótesis nula.

t en 4 (2,45) no es más extremo que el t de corte en 3 (4,541).

Por lo tanto, no se rechaza la hipótesis nula.

b)



c) $d = M/S = 5/\sqrt{16{,}67} = 5/4{,}08 = 1{,}23$

d) Explicación: lo primero que hice fue simplificar las cosas convirtiendo los números en "valores de cambio" para cada ciudad, cantidad de residuos con posterioridad al programa (1998) menos cantidad de residuos con anterioridad al programa (1997). Después calculé la media de esos valores de cambio, que era 5, y que indica una disminución de cinco libras de residuos por manzana por día.

El siguiente paso fue analizar si este resultado, correspondiente a estas cinco ciudades, indica alguna diferencia real en forma general como consecuencia del programa. La otra alternativa es la posibilidad de que este cambio podría haber ocurrido en cualquiera de las otras cuatro ciudades seleccionadas al azar sólo por casualidad, aun cuando el programa en general no tuviera ningún efecto real. Es decir, imaginamos que el cambio promedio en ciudades en general que implementan este programa es realmente 0, y tal vez sólo ocurrió que para este estudio se seleccionaron cuatro ciudades que, de todos modos, hubieran disminuido su nivel de residuos.

Entonces, determiné precisamente cuánto tendría que cambiar un grupo de cuatro ciudades antes de que yo pudiera concluir que han cambiado demasiado como para atribuírselo a la casualidad. Esto requirió el cálculo de las características de esa población hipotética de ciudades en la que, en promedio, "no existe ningún cambio". Su media sería 0 cambio (es decir, una media de cambio igual a 0 es exactamente como describiríamos un promedio de ausencia de cambio). Dado que no conocía la varianza de esa distribución hipotética de ciudades en las que no se producía ningún cambio, la estimé a partir de la información proveniente de la muestra de las cuatro ciudades. Si la muestra era sólo una elección casual de la población hipotética, su varianza debería ser representativa de la población hipotética.

Sin embargo, la varianza calculada a partir de la muestra en general será levemente menor que la verdadera varianza poblacional, ya que la varianza de la muestra se basa en desvíos cuadráticos de la media muestral. La media de la muestra es el punto de equilibrio perfecto de sus valores, de manera que la suma de los desvíos cuadráticos calculada a partir de ella será menor que la suma de desvíos cuadráticos calculada a partir de cualquier otro número (tal como la media poblacional). Por lo tanto, tuve que modificar la fórmula de la varianza para tener en cuenta esa diferencia: en lugar de dividir la suma de los desvíos cuadráticos por la cantidad de valores, la dividí por los "grados de libertad", que es la cantidad de valores menos 1, en este caso 3. (El ajuste tiene en cuenta precisamente la tendencia de la varianza de una muestra a subestimar la verdadera varianza poblacional). Tal como lo indican los cálculos en los pasos de la prueba de hipótesis, esto dio una varianza poblacional estimada (S^2) igual a 16,67.

Como no estaba interesado en ciudades individuales sino en un grupo de cuatro, lo que realmente necesitaba saber eran las características de una distribución de todas las medias posibles de muestras formadas por cuatro ciudades, tomadas de esa población hipotética de valores de cambio correspondientes a ciudades individuales. Tal distribución de medias tendrá la misma media (ya que no existe razón para esperar que las medias de esos grupos de cuatro tomados al azar sean sistemáticamente mayores o menores que 0). Pero esa distribución tendrá una varianza mucho menor (porque es mucho menos probable que sea extremo el promedio de un grupo de cuatro valores que cualquier valor individual). Afortunadamente, se sabe (y se puede probar matemáticamente) que la varianza de una distribución de medias es la varianza de la distribución de observaciones individuales dividida por la cantidad de individuos de cada muestra. En nuestro ejemplo, esto es igual a 16,67 dividido 4, lo que da 4,17. Por lo tanto, el desvío estándar de esta distribución es la raíz cuadrada de 4,17, es decir, 2,04.

También ocurre que si podemos suponer que la población hipotética de valores de cambio de las ciudades individuales está normalmente distribuida (y no tenemos razón para pensar de otro modo), podemos considerar que la distribución de medias de muestras de esa distribución tiene una forma precisa conocida, denominada distribución *t* (que tiene colas levemente más altas que la curva normal). Si buscamos en una tabla para distribución *t* el caso en el que hay 3 grados de libertad utilizados para estimar la varianza poblacional, la tabla indica que existe menos de un 1% de probabilidad de obtener un valor que se encuentre a una distancia de 4,541 desvíos estándar o más de la media de esa distribución.

El valor de cambio medio de la presente muestra de cuatro ciudades era 5, lo que daría 2,45 (es decir, 5/2,04) desvíos estándar por encima de la media de 0 cambio, en la distribución de medias de registros de cambio. Dado que este resultado no es tan extremo como 4,541, existe más de un 1% de probabilidad de que esos resultados pudieran haber surgido de una distribución hipotética en la que no se producía ningún cambio. Por lo tanto, el investigador no puede descartar esa posibilidad, y se diría que el experimento no es concluyente.

Finalmente, es posible describir el nivel del efecto en un formato estandarizado denominado tamaño del efecto *(d)*. El tamaño del efecto es precisamente la media de los valores de cambio dividida por el desvío estándar poblacional estimado, en este caso, 5 dividido 4,08, lo que da 1,23. Esto significa que el cambio entre antes y después del programa fue de más de 1 desvío estándar; en consecuencia, se trata de un cambio bastante considerable. Sin embargo, aun con ese importante nivel de cambio, el resultado no fue significativo (indudablemente debido al tamaño tan pequeño de la muestra, que estaba formada sólo por cuatro ciudades).

5. Según tabla 9–9:
   a) 0,22; b) 0,71; c) 0,86; d) 0,77; e) 0,99.
6. La media es el promedio ordinario (la suma de los valores dividida por la cantidad de valores). Por lo tanto, la primera parte de este resultado indica que los valores promedio bajo luz brillante fueron ligeramente mayores que bajo luz tenue. Pero lo importante es la última parte. Al decir que la diferencia "no fue significativa", el investigador está indicándonos que este pequeño grado de diferencia podría haber sido encontrado fácilmente entre los 20 individuos probados, aun si las personas en general no difieren bajo condiciones de luz brillante o tenue. La última parte, "$t(19) = 1,62$", se refiere a los detalles de cómo se determinó que la diferencia "no era significativa".

   La lógica implícita depende de imaginar primero una distribución hipotética de diferencias en condiciones de luz brillante y tenue en la cual la diferencia promedio es 0. Esta distribución también tendrá una cantidad específica de variación que el investigador debe estimar sobre la base de la variación de las diferencias de las 20 personas, obtenida en el experimento. La fórmula para estimar esta variación requiere tomar la diferencia de cada persona menos la media de todas las diferencias. Luego, cada uno de estos "desvíos" son elevados al cuadrado y sumados. Dividiendo esta suma por la cantidad de participantes (20 en este caso) obtenemos la "varianza". Sin embargo, para estimar la varianza del grupo hipotético mayor, deben realizarse ciertos ajustes. La suma de los desvíos cuadráticos no se divide por la cantidad de participantes sino por la cantidad de casos menos 1 (19 en este estudio). (El ajuste es necesario porque la variación de las personas en general, estimada a partir del grupo particular bajo estudio, será demasiado pequeña porque los desvíos cuadráticos son desvíos del promedio del grupo estudiado en particular, hecho que tiene el efecto de minimizar el total).

De todos modos, lo que realmente se necesita es imaginar una distribución hipotética formada por los promedios de las diferencias de grupos de 20 personas, las 20 diferencias que forman cada promedio surgen de la primera distribución hipotética de diferencias de individuos, mencionada con anterioridad, y que pertenece a un mundo en el que la diferencia promedio general es 0. Esta nueva distribución hipotética formada por promedios de 20 diferencias también tendrá un promedio de 0, pero su varianza será mucho menor porque es menos probable que tal distribución de promedios tenga diferencias extremas. Su varianza, de hecho, resulta ser la varianza de la primera distribución hipotética dividida por la cantidad de participantes de cada grupo (en este caso 20).

Esta distribución de promedios de grupos de 20 diferencias también tendrá una forma conocida, denominada "distribución t". (En realidad, no necesariamente debe tener esa forma, pero dado que el investigador utilizó la t en la descripción, debe haber supuesto que se daban las condiciones adecuadas). Existen diferentes distribuciones t de acuerdo con el número que se utilizó para realizar la división para estimar la varianza de la distribución hipotética de diferencias individuales, que en nuestro caso fue 19. (De aquí es de donde proviene el 19 en el paréntesis).

Finalmente, podemos buscar en una tabla cuán alto debería ser un promedio de 20 diferencias para ubicarse dentro del 5% más alto de esa distribución t. Ese número es 1,729 desvíos estándar del promedio de esa distribución de promedios (un desvío estándar es la raíz cuadrada de la varianza, es una medida estándar de variación). En este estudio en particular, sin embargo, el investigador nos ha indicado que la cantidad de desvíos estándar, por encima de la media de la distribución hipotética en que se ubicará el promedio de los 20 diferencias reales, era sólo 1,62 (este es el punto t). Dado que este número no se ubica dentro del 5% superior (el intervalo que comienza con 1,729), el investigador no puede descartar la posibilidad de que este grupo de 20 podría haber salido de la distribución hipotética en la que la diferencia promedio es, de hecho, 0.

Es norma convencional en psicología que un resultado sea considerado "significativo" sólo si la posibilidad de que surja de una situación en la que realmente no existe diferencia es menor al 5%. Cuando un resultado no es significativo, el estudio no es concluyente.

7. Angustia: $S_M = \sqrt{S^2/N} = 1{,}852/100 = \sqrt{0{,}034} = 0{,}185$; $t = 1{,}50/0{,}185 = 8{,}11$.
Depresión: $S_M = \sqrt{(4{,}23)2/100} = 0{,}423$; $t = 3{,}08/0{,}423 = 7{,}28$.
Introversión: $S_M = 0{,}222$; $t = 0{,}23/0{,}222 = 1{,}04$.
Neurotismo: $S_M = 0{,}421$; $t = 0{,}89/0{,}421 = 2{,}11$.
La explicación de la prueba t es básicamente la misma que la del ejercicio 6.

Capítulo 10

1. a) Una prueba t para medias dependientes se utiliza cuando cada participante es probado bajo dos condiciones (tales como antes y después de algún tratamiento), de manera que hayan dos valores observados por participante. Una prueba t para medias independientes se utiliza cuando algunos participantes son probados una vez bajo una de las condiciones y otros son probados una vez bajo otra condición diferente, de forma tal que haya sólo un valor por participante.
2. i) t necesario ($gl = 58$, $p < 0{,}05$, dos colas) $= \pm 2{,}004$; $S^2_{\text{Combinada}} = [gl_1/(gl_1 + gl_2)](S^2_1) + [gl_2/(gl_1 + gl_2)](S^2_2) = (29/58)(2{,}4) + (29/58)(2{,}8) = 1{,}2 + 1{,}4$; $S^2_{M1} = S^2_{\text{Combinada}}/N_1 = 2{,}6/30 = 0{,}087$; $S^2_{M2} = 0{,}087$; $S^2_{\text{Diferencia}} = S^2_{M1} + S^2_{M2} = 0{,}174$; $S_{\text{Diferencia}} = \sqrt{S^2_{\text{Diferencia}}} = \sqrt{174} = 0{,}417$; $t = (M_1 - M_2)/S_{\text{Diferencia}} = (12 - 1{,}11)/0{,}417 = 2{,}16$. Conclusión: se rechaza la hipótesis nula. La diferencia es significativa. Tamaño del efecto:
$d = (M - M_2)/S_{\text{Combinada}} = (12 - 11{,}1)/\sqrt{2{,}6} = 0{,}9/1{,}6 = 0{,}56$ (tamaño del efecto aproximadamente mediano).. Potencia (de la tabla) $= 0{,}47$.
   ii) t necesario ($gl = 58$, $p < 0{,}05$, dos colas) $= \pm 2{,}004$; $S^2_{\text{Combinada}} = [gl_1/(gl_1 + gl_2)](S^2_1) + [gl_2/(gl_1 + gl_2)](S^2_2) = (19/58)(2{,}4) + (39/58)(2{,}8) = (0{,}328)(2{,}4) + (0{,}672)(2{,}8) = 0{,}787 + 1{,}882 = 2{,}7$; $S^2_{M1} = S^2_{\text{Combinada}}/N_1 = 2{,}7/20 = 0{,}135$; $S^2_{M2} = 2{,}7/40 = 0{,}068$; $S^2_{\text{Diferencia}} = S^2_{M1} + S^2_{M2} = 0{,}203$; $S_{\text{Diferencia}} = \sqrt{S^2_{\text{Diferencia}}} = \sqrt{0{,}203} = 0{,}451$; $t = (M_1 - M_2)/S_{\text{Diferencia}} = (12 - 11{,}1)/0{,}451 = 0{,}9/0{,}451 = 2{,}00$.
Conclusión: no se rechaza la hipótesis nula. La diferencia no es significativa. Tamaño del efecto: $d = 0{,}9/\sqrt{2{,}7} = 0{,}55$ (tamaño del efec-

to aproximadamente mediano). Potencia: $N' = [(2)(20)(40)]/(20 + 40) = 26,7$; potencia (de la tabla) está entre 0,33 y 0,47.

iii) t necesario ($gl = 58$, $p < 0,05$, dos colas) $= \pm 2,004$; $S^2_{\text{Combinada}} = 2,6$; $S^2_{M1} = 0,087$; $S^2_{M2} = 0,087$; $S_{\text{Diferencia}} = 0,417$; $t = 2,16$. Conclusión: se rechaza la hipótesis nula. La diferencia es significativa. Tamaño del efecto: $d = 0,9/\sqrt{2,6} = 0,56$ (tamaño del efecto aproximadamente mediano). Potencia $= 0,47$.

3. a) Pasos de la prueba de hipótesis:

1. Replantear el problema en función de hipótesis de investigación e hipótesis nula de poblaciones.

Población 1: personas que se informan a través de la TV.

Población 2: personas que se informan a través de la radio.

La hipótesis de investigación establece que las dos poblaciones tienen medias diferentes. La hipótesis nula establece que las dos poblaciones tienen la misma media.

2. Determinar las características de la distribución comparativa.

Varianza poblacional estimada $= S^2_{\text{Combinada}} = (60/80)(4) + (20/80)(6) = 3,0 + 1,5 = 4,5$.

Distribución comparativa (distribución de diferencias de medias): Media $= 0$; $S_{\text{Diferencia}} = 0,54$; Forma $= t(80)$.

Cálculo de $S_{\text{Diferencia}}$: $S^2_{M1} = 4,5/61 = 0,074$; $S^2_{M2} = 4,5/21 = 0,214$; $S^2_{\text{Diferencia}} = 0,074 + 0,214 = 0,288$; $S_{\text{Diferencia}} = 0,54$.

3. Determinar el punto de corte en la distribución comparativa, en el que debería rechazarse la hipótesis nula.

t necesario ($gl = 80$, $p < 0,01$, dos colas) $= \pm 2,639$.

4. Determinar el valor muestral en la distribución comparativa.

$t = (24 - 26)/0,54 = -2/0,54 = -3,70$.

5. Comparar los valores de los pasos 3 y 4 para decidir si se rechaza o no la hipótesis nula.

t del paso 4 ($-3,70$) es más extremo que el punto t de corte del paso 3 ($\pm 2,639$). Por lo tanto, se rechaza la hipótesis nula; se llega a la conclusión de que la predicción es sustentada por el experimento.

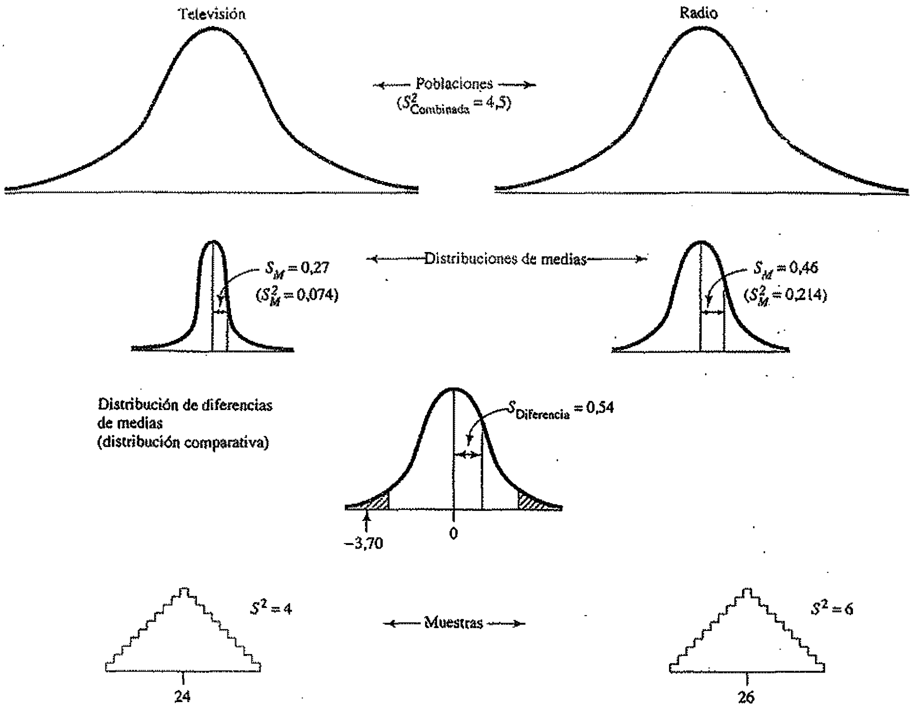
b) $d = (24 - 26)/\sqrt{4,5} = -2/2,12 = -0,94$; gran tamaño del efecto.

c) *Véase* figura en la parte superior de la próxima página.

d) Explicación: la media (M) es el promedio aritmético (la suma de los valores observados dividida por la cantidad de ellos). En este caso, el grupo que se informaba por radio tenía un valor promedio más alto en la prueba que el grupo que se informaba por TV. S^2 se refiere a la estimación de la varianza de los valores de la población general basada en la varianza de los valores del grupo de personas bajo estudio (denominada muestra). La varianza (S^2) es una medida del grado de variación en un grupo de valores. Al estimar la varianza de la población, a partir de la varianza de la muestra, la diferencia de cada valor con respecto a la media se eleva al cuadrado y la suma de esas diferencias cuadráticas se divide por los grados de libertad, es decir, por la cantidad de participantes de la muestra menos 1. (Los grados de libertad representan la cantidad de información única disponible en la muestra para utilizar en la estimación de la población. Utilizar la varianza de la muestra, que es la suma de las diferencias cuadráticas dividida por la cantidad de casos, daría una estimación demasiado pequeña de la varianza poblacional). En este caso, se obtienen dos estimaciones, una de cada muestra.

Ahora que ya hemos analizado los resultados del problema, veamos de qué modo hemos llegado a las conclusiones. La manera de formular el problema es preguntarse cuál es la probabilidad de obtener esta diferencia entre los dos grupos con respecto a los valores sobre conocimiento, si no hubiera diferencia entre la radio y la TV. Es decir, si los grupos que se informan a través de la TV y la radio realmente representaban dos poblaciones mayores que no eran diferentes entre sí, ¿cuál es la probabilidad de que pudieran haber obtenido una muestra de cada población que fueran tan diferentes entre sí?

Para contestar esta pregunta es necesario calcular cómo se verían tales poblaciones no diferentes entre sí. Aquí se aplican las estimaciones de la varianza de población. De hecho, aun si los dos grupos representaran poblaciones diferentes (sólo las medias serían diferentes), se presume que la varianza es la misma. Por lo tanto, se trata de dos estimaciones de la misma varianza de población, y podemos promediar ambas estimaciones para obtener una estimación aún mejor. Sin embargo, al promediar es necesario dar mayor peso a la estimación basada en mayores grados de libertad. Entonces, se calcula un promedio ponderado multiplicando cada estimación por su proporción en los grados totales de libertad y sumando los resultados. Esta estimación combinada



de la varianza poblacional es igual a 4,5. Hasta aquí se había estimado sólo la varianza de la población de valores sobre conocimientos correspondientes a los individuos.

Ahora bien, dado que lo que nos interesa no eran los valores individuales sino la diferencia entre la media de un grupo de 61 y la media de otro grupo de 21, necesitábamos calcular cuáles serían las características de una distribución de todas las posibles diferencias de medias de grupos de 60 y 21 tomados al azar de las dos poblaciones idénticas, cuyas varianzas acabamos de estimar. Este cálculo requería de dos pasos:

Primero, necesitábamos calcular las características de la distribución intermedia de cada muestra, es decir, de la distribución de medias de todas las muestras posibles de ese tamaño tomadas de esa población. Para el grupo de la TV esta sería una distribución de medias muestrales de 61 valores cada una. Tal distribución tendrá una varianza mucho menor que la varianza de la población de observaciones individuales de donde provienen las muestras, ya que cualquier media tiene menos probabilidades de ser extrema que cualquier valor individual (porque la media de varios valores probablemente incluya algunos valores que equilibran o reducen el efecto de cualquier extremo). De hecho, se puede demostrar matemáticamente que la varianza de una distribución de medias de todas las muestras posibles será exactamente la varianza de la población de observaciones individuales de origen dividida por la cantidad de observaciones en cada muestra. Para el grupo de TV, esta distribución sería 4,5 dividido 61, es decir, 0,074. La cifra correspondiente para el grupo de la radio es 0,214.

El segundo paso se refiere directamente a la distribución de diferencias de medias. Es la distribución que surgiría si tomáramos una media de la distribución de medias de todas las muestras posibles del grupo de TV, y tomáramos otra media de la distribución semejante correspondiente al grupo de la radio y calculáramos la diferencia. Después de hacer esto muchas veces, la distribución de diferencias obtenida del modo descripto crearía una nueva *distribución de diferencias de medias*. Ya que suponemos (si no había diferencia entre radio y TV) que las dos poblaciones originales tenían las mismas medias, las dos distribuciones de muchas medias de muestras deberían tener la misma media también. En promedio, la diferencia entre una media tomada del grupo de la TV y una media tomada del grupo de la radio debería dar 0 (porque algunas veces será mayor una y otras veces la otra, pero a la larga estas fluctuaciones aleatorias deberían equilibrarse). La varianza de la distribución de diferencias de medias será afectada por la variación en ambas distribuciones de medias; de hecho, será simplemente igual a la suma de las dos. Por lo tanto, la varianza será 0,074 más 0,214, lo que da 0,288. En realidad, la variación en tales distribuciones se describe más frecuentemente en términos de lo que se denomina desvío estándar (la raíz cuadrada de la varianza), que en este caso es la raíz cuadrada de 0,288, o lo que es lo mismo, 0,54.

También resulta que estas distribuciones de diferencias de medias tienen una forma precisa conocida, por lo tanto, es posible buscar en una tabla la probabilidad de estar a una cierta distancia más allá de su media. La distancia se mide en desvíos estándar. En este caso, la tabla indica que en la distribución (con un total de 80 grados de libertad) existe menos del 1% de probabilidad de obtener un valor (una diferencia de medias) que se ubique a 2,639 o más desvíos estándar de la media en cualquier dirección. (Tuvimos en cuenta ambas direcciones porque estábamos analizando si existía una diferencia en cualquier dirección entre los grupos de TV y de radio. El "nivel 1%" se refiere a la puntuación convencional en la cual los científicos sociales, que se preocupan mucho por no correr el riesgo de concluir erróneamente que un experimento ha dado una diferencia, deciden que es demasiado improbable que haya sucedido algo por casualidad). Hemos representado gráficamente las distintas distribuciones relacionadas con este ejercicio (*véase* la figura de la página anterior). *Véase* dónde quedó impresa la figura a la que hace referencia.

La diferencia entre las dos medias en particular fue –2 (es decir, 24 – 26). La diferencia sería igual a 3,70 (es decir, 2/0,54) desvíos estándar por debajo de la media en la distribución de todas las diferencias posibles de medias de grupos de este tamaño. Dado que el resultado es más extremo que –2,639, se podría rechazar por improbable la posibilidad de obtener una diferencia de este tamaño, tomando al azar dos grupos cualesquiera de participantes independientemente de si se habían estado informando a través de la TV o la radio. Por lo tanto, el investigador puede considerar los resultados de este estudio como soporte de su predicción.

Más aún, el investigador deseaba saber no sólo que los resultados no eran casuales sino también cuál era el tamaño del efecto producido por informarse a través de la radio o de la TV. *La diferencia entre las dos medias era de* 2 puntos en la medida de conocimiento. La cantidad típica de variación de valores en cualquier escala se describe a través del desvío estándar (la raíz cuadrada de la varianza, siendo la varianza el promedio de los cuadrados de la diferencia de cada valor con respecto a la media). En este caso, el desvío estándar que estimaríamos utiliza información de ambas muestras; es una estimación combinada. La estimación combinada de la varianza era 4,5, su raíz cuadrada es 2,12. Por lo tanto, una diferencia de 2 puntos en la escala es una diferencia de casi 1 desvío estándar (0,94 desvíos estándar). En las investigaciones sociales en general, un tamaño del efecto de 0,80 se considera grande, por lo que podemos decir que claramente se trata de un gran efecto. Por lo tanto, además de la conclusión de que no es probable que el resultado haya surgido sólo por casualidad, el investigador también puede concluir que la ventaja de la radio por sobre la TV es bastante considerable.

4. a) Los cinco pasos de la prueba de hipótesis deberían presentarse en forma paralela a los del ejercicio 3a. La información clave para este problema es la siguiente:

t necesario (gl = 9, dos colas, $p < 0{,}05$) = ±2,262.

Normales: $M = 36/6 = 6$; $S^2 = 28/5 = 5{,}6$,

Propio nombre: $M = 48/5 = 9{,}6$; $S^2 = 77{,}2/4 = 19{,}3$.

$S^2_{\text{Combinada}} = (5/9)(5,6) + (4/9)(19,3) = 3,11 + 8,58 = 11,69$.

$S^2_{M1} = 11,69/6 = 1,95$; $S^2_{M2} = 11,69/5 = 2,34$.

$S^2_{\text{Diferencia}} = 1,95 + 2,34 = 4,29$; $S_{\text{Diferencia}} = 2,07$

$t = (6 - 9,6)/2,07 = -3,6/2,07 = -1,73$.

No se rechaza la hipótesis nula; el experimento no es concluyente en cuanto a si el hecho de incluir el nombre del niño produce alguna diferencia.

Nota: el problema, en realidad, tiene un defecto en el sentido de que, aparentemente, no cumple con el supuesto que requiere iguales varianzas poblacionales. Sin embargo, dado que el resultado no fue significativo aun utilizando el procedimiento ordinario, podemos presumir que probablemente no habría sido significativo utilizando un procedimiento modificado.

b) $d = (6 - 9,6) / \sqrt{11,69} = -3,6/3,42 = 1,05$.

c) La ilustración gráfica sería similar a la indicada en la respuesta al ejercicio 3c.

d) La descripción para una persona que no ha estudiado la prueba t para medias independientes sería similar a la de la respuesta al ejercicio 3d, excepto que no tendríamos que explicar todos los puntos que la persona ya conoce si comprende la prueba t para medias dependientes.

5. La respuesta debería explicar lo siguiente, pero debería estar redactada de manera tal de explicar todos los términos y conceptos (como en la respuesta al ejercicio 3, por ejemplo).

El estudio refleja que utilizando un nivel de significación convencional de 0,05, los niños alemanes que reciben bajos niveles de apoyo –ya sea de su madre, su padre o sus compañeros–, mostraban menores niveles de auto-valoración. Más aún, los tamaño del efecto eran bastante grandes ($d = 0,78$ y $d = 0,69$) con respecto al apoyo de la madre o del padre; no obstante, el tamaño del efecto era sólo entre pequeño y moderado ($d = 0,35$) con respecto al apoyo de los compañeros. Lo anterior parecería implicar que el apoyo de los padres es más importante que el apoyo de los compañeros en cuanto al sentimiento de autovaloración del niño. La potencia del estudio para un gran tamaño del efecto es 0,98. (Presumiendo que había aproximadamente igual cantidad de niños en los dos grupos (el de nivel de apoyo alto y el de nivel de apoyo bajo), que la prueba es de dos colas y que se utiliza una cantidad de 50 observaciones en cada grupo). La potencia para un tamaño del efecto mediano es 0,70. Debido a que ya sabemos que los resultados son significativos y conocemos los tamaños del efecto, los cálculos de la potencia no son muy importantes.

6. a) $d = (107 - 149)/84 = -42/84 = -0,50$. Tamaño del efecto mediano. Cantidad necesaria de participantes por grupo para un tamaño del efecto mediano, $p < 0,05$, una cola (de la tabla 10-7) = 50; 100 participantes en total.

b) $d = (22,5 - 16,2)/31,5 = 0,20$. Tamaño del efecto pequeño. N necesaria: 393 por grupo, 786 en total.

c) $d = (14 - 12)/2,5 = 0,80$. Gran tamaño del efecto. N necesaria: 20 por grupo, 40 en total.

d) $d = (480 - 520)/50 = -0,80$. Gran tamaño del efecto. N necesaria: 26 por grupo, 52 en total.

Capítulo 11

1. a) F necesario ($gl = 2, 27$; $p < 0,05$) = 3,36; $S^2_{\text{Entre}} = (Sc/gl)(n) = \{[(7,4 - 7)^2 + (6,8 - 7)2 + (6,8 - 7)^2]/\_(3 - 1)\}(10) = (0,24/2)(10) = 1,2$; $S^2_{\text{Dentro}} = (0,82 + 0,90 + 0,80)/3 = 0,84$; $F = 1,2/0,84 = 1,43$; no se rechaza la hipótesis nula; los grupos no son significativamente diferentes al nivel 0,05. Tamaño del efecto: $f = \sqrt{1,43} / \sqrt{10} = 1,20/3,16 = 0,40$ (gran tamaño del efecto). Potencia = 0,45.

b) F necesario ($gl = 3, 96$; $p < 0,05$) = 2,70 (en realidad utilizando $gl = 3, 95$); $S^2_{\text{entre}} = (164,67)(25) = 4.116,75$; $S^2_{\text{dentro}} = (242 + 282 + 312 + 252)/4 = 736,5$; $F = 4.116,75/736,5 = 5,59$; se rechaza la hipótesis nula, los grupos son significativamente diferentes al nivel 0,05. Tamaño del efecto: $f = \sqrt{5,59} / \sqrt{25} = 0,47$ (gran tamaño del efecto). Potencia entre 0,85 y 0,96.

c) F necesario ($gl = 4, 120$; $p < 0,05$) = 2,46 (en realidad utilizando $gl = 4, 100$); $S^2_{\text{entre}} = (123,5)(25) = 3.087,5$; $S^2_{\text{dentro}} = (242 + 282 + 312 + 252 + 272)/5 = 735$; $F = 3.087,5/735 = 420$; se rechaza la hipótesis nula; los grupos son significativamente diferentes al nivel 0,05. Tamaño del efecto: $f = \sqrt{4,20}/\sqrt{25} = 0,41$ (gran tamaño de efecto). Potencia entre 0,90 y 0,98.

2. a) F necesario ($gl = 2, 9$; $p < 0,01$) = 8,02.
Grupo 1: $M = 8$, $S^2 = 0,67$; Grupo 2: $M = 6$, $S^2 = 0,67$; Grupo 3: $M = 4$, $S^2 = 0,67$.
$S^2_{\text{entre}} = (4)(4) = 16$; $S^2_{\text{dentro}} = 0,67$; $F = 16/0,67 = 23,88$; se rechaza la hipótesis nula; los grupos son significativamente diferentes al nivel 0,01.

b) F necesario ($gl = 2, 9$; $p < 0,01$) = 8,02.
Grupo 1: $M = 8$, $S^2 = 21,33$; Grupo 2: $M = 6$, $S^2 = 21,33$; Grupo 3: $M = 4$, $S^2 = 21,33$.

$S^2_{entre} = (4)(4) = 16$; $S^2_{dentro} = 21{,}33$; $F = 16/21{,}33 = 0{,}75$; no se rechaza la hipótesis nula; los grupos no son significativamente diferentes al nivel 0,01.

3. a) Pasos de la prueba de hipótesis:

1. Replantear el problema en función de hipótesis de investigación e hipótesis nula de las poblaciones.

Población 1: pacientes con trastornos afectivos.

Población 2: pacientes con trastornos cognitivos.

Población 3: pacientes con trastornos relacionados con las drogas.

La hipótesis de investigación establece que las tres medias poblacionales son diferentes. La hipótesis nula establece que las tres poblaciones tienen la misma media.

2. Determinar las características de la distribución comparativa.

Distribución F con 2 y 9 grados de libertad.

3. Determinar el punto de corte en la distribución comparativa, a partir del cual se debería rechazar la hipótesis nula.

Nivel 5%, $F(2,9)$ necesario $= 4{,}26$.

4. Determinar el valor muestral en la distribución comparativa.

Estimación intragrupal de la varianza poblacional $(S^2_{dentro}) = (0{,}67 + 3{,}33 + 2{,}67) / 3 = 2{,}22$

Estimación intergrupal de la varianza poblacional $(S^2_{entre}) = (5{,}33)(4) = 21{,}32$

Razón $F = 21{,}32/2{,}22 = 9{,}60$.

5. Comparar los valores obtenidos en los pasos 3 y 4 para decidir si se rechaza o no la hipótesis nula.

F del paso 4 (9,60) es más extremo que el corte F del paso 3 (4,26).

Por lo tanto, se rechaza la hipótesis nula; se sostiene la hipótesis de investigación; existe una diferencia significativa.

b) Explicación: la hipótesis nula establece que los tres grupos representan poblaciones formadas por los tiempos de internación, en semanas, que tienen las mismas medias (y, al igual que con la prueba t, debemos estar en condiciones de presumir que tienen iguales varianzas). Si ésta hipótesis nula es verdadera, entonces podemos estimar la varianza de esas poblaciones iguales de dos modos:

1. Podemos estimarla a partir de la variación dentro de cada uno de los tres grupos y luego promediarlos (es exactamente lo que haríamos en una prueba t para medias independientes, excepto que ahora se están promediando tres grupos en lugar de promediar sólo dos; además, en una prueba t ponderaríamos estas varianzas según los grados de libertad con los que cada una contribuye a la estimación total. Sin embargo, dado que los tres grupos tienen la misma cantidad de observaciones, podemos simplemente promediarlos; en efecto, sería lo mismo que ponderarlos de manera pareja). En este ejemplo, las tres estimaciones de varianza eran 0,67, 3,33 y 2,67, lo que dio una estimación combinada de 2,22. A esto se denomina estimación intragrupal de la varianza poblacional.

2. Podemos estimar la varianza utilizando las tres medias. Si presumimos que la hipótesis nula es verdadera, las medias de los tres grupos se basan en muestras tomadas de poblaciones idénticas. Cada una de estas poblaciones idénticas tendrá una distribución de medias muestrales idéntica a las demás, tomada de esa población. Las medias de las tres muestras provienen todas de poblaciones idénticas, es decir, que es lo mismo que si pertenecieran todas a la misma población. Por eso, la cantidad de variación entre las tres medias debería ser representativa de la variación en la distribución de medias de donde puede considerarse que provienen. Por consiguiente, podemos utilizar estas tres medias (6, 10 y 10) para estimar la varianza de esa distribución de medias. Utilizando la fórmula usual para estimar una varianza poblacional, obtenemos 5,33.

Sin embargo, lo que necesitamos es una distribución de observaciones individuales. Por lo tanto, la cuestión es la siguiente: ¿Cuál sería la distribución de observaciones que produciría una distribución de medias (de cuatro registros cada una) con una varianza de 5,33? Para encontrar la distribución de medias de una distribución de observaciones individuales, dividimos la varianza de la distribución de observaciones por el tamaño de las muestras. En este caso, queremos hacer lo contrario. En consecuencia, multiplicamos la varianza de la distribución de medias por el tamaño de las muestras para obtener la varianza de la distribución de individuos. El resultado es igual a 5,33 por 4, es decir, 21,32. A esto se denomina estimación intergrupal de la varianza poblacional.

Si la hipótesis nula es verdadera, las dos estimaciones deberían ser aproximadamente iguales porque estiman esencialmente la misma población. Por lo tanto, la razón resultante de dividir la estimación intergrupal por la estimación intragrupal debería ser aproximadamente 1.

Pero si la hipótesis nula es falsa y las tres poblaciones que representan estos grupos tienen diferentes medias, la estimación basada en la variación entre las medias de grupos será mayor que la que se basa en al variación dentro de los grupos. La razón por la que será mayor es la siguiente: si la hipótesis nula es verdadera, la única razón para que las medias de los grupos varíen es por la varianza dentro de cada una de las tres distribuciones idénticas de medias. Pero si la hipótesis nula es falsa, cada una de esas distribuciones de medias también tiene una media diferente. Por lo tanto, la variación en las medias se debe a la variación dentro de cada una de esas distribuciones de medias, en este caso no idénticas, pero **también** a las diferencias entre las medias de esas distribuciones de medias. En suma, existe una fuente adicional de variación en las medias de los grupos. Si estimamos la varianza poblacional utilizando esas tres medias, la estimación será mayor de lo que debería si la hipótesis nula fuera verdadera. Por otro lado, la varianza intragrupal no se ve afectada por el hecho de que los tres grupos tengan diferentes medias, porque sólo tiene en cuenta la variación dentro de cada uno de los grupos. La varianza intragrupal, por lo tanto, no aumenta para nada si la hipótesis nula es falsa. Por eso, cuando la hipótesis nula es falsa, la razón entre la varianza intergrupal y la varianza intragrupal será mayor que 1.

La razón entre la estimación intergrupal y la estimación intragrupal se denomina razón F. En este ejemplo, nuestra razón F es 21,32 a 2,22: 21,32/2,22 = 9,60.

Los estadísticos han construido tablas que indican lo que sucede cuando calculamos razón F habiendo seleccionado un grupo de cuatro valores al azar de cada una de tres poblaciones idénticas. Esa es precisamente la situación en la que nuestra hipótesis nula es verdadera. Buscando en esas tablas, descubrimos que existe menos de un 5% de posibilidad de obtener una razón F mayor a 4,26, y dado que la razón F real es mayor a ese número, podemos rechazar la hipótesis nula.

4. F necesario ($gl = 2, 147$; $p < 0{,}05$) = 3,09 (en realidad, utilizando $gl = 2, 100$); $S^2_{entre} = (0{,}09)(50) = 4{,}5$; $S^2_{dentro} = (5{,}2 + 5{,}8 + 4{,}8)/3 = 5{,}27$; $F = 4{,}5/5{,}27 = 0{,}85$; no se rechaza la hipótesis nula; los grupos no son significativamente diferentes al nivel 0,05. Tamaño del efecto: $f = \sqrt{0{,}85}/\sqrt{50} = 0{,}13$ (pequeño tamaño del efecto). Potencia = 0,18.
Con respecto a los cinco pasos de la prueba de hipótesis y la explicación, ***véase*** la respuesta al ejercicio 3; también se utiliza material de las respuestas a los ejercicios acerca de la prueba t del capítulo 10.
5. El resultado sostiene la hipótesis que establece que los reclusos en los tres tipos de prisión tienen diferentes grados de necesidad en cuanto al cuidado de su salud mental. El tamaño del efecto es 0,38. (La explicación completa incluiría el material de la respuesta al ejercicio 3, más material de las respuestas a los ejercicios acerca de la prueba t del capítulo 10).

Capítulo 12

1. $gl_{Total} = N - 1 = 12 - 1 = 11$.
$gl_{dentro} = gl^1 + gl^2 + \ldots + gl_{Último}$
$= (4-1) + (4-1) + (4-1) = 3 + 3 + 3 = 9$.
$gl_{entre} = N_{Grupos} - 1 = 3 - 1 = 2$.
F necesario para $gl = 2, 9$ al nivel 0,01 = 8,02.

Grupo 1

X	$X - GM$		$X - M$		$M - GM$	
	Desv	*Desv*²	*Desv*	*Desv*²	*Desv*	*Desv*²
8	2	4	0	0	2	4
8	2	4	0	0	2	4
7	1	1	−1	1	2	4
9	3	9	1	1	2	4
Σ 32		18		2		16

$M = 32/4 = 8$.

Grupo 2

X	$X - GM$		$X - M$		$M - GM$	
	Desv	*Desv*²	*Dessv*	*Desv*²	*Desv*	*Desv*²
6	0	0	0	0	0	0
6	0	0	0	0	0	0
5	1	−1	−1	1	0	0
7	1	1	1	1	0	0
Σ 24		2		2		2

$M = 24/4 = 6$.

Grupo 3

X	$X-GM$		$X-M$		$M-GM$	
	Desv	Desv²	Desv	Desv²	Desv	Desv²
4	−2	4	0	0	−2	4
4	−2	4	0	0	−2	4
3	−3	9	−1	1	−2	4
5	−1	1	1	1	−2	4
Σ 16		18		2		16

$M = 16/4 = 6$.

$GM = (32 + 24 + 16)/12 = 72/12 = 6$.

$SC_{Total} = 18 + 2 + 18 = 38$.
$SC_{dentro} = 2 + 2 + 2 = 6$.
$SC_{entre} = 16 + 0 + 16 = 32$.

Tabla de análisis de varianza:

Fuente	SC	gl	CM	F
Intergrupal	32	2	16	23,88
Intragrupal	6	9	0,67	
Total		38	11	

Conclusión: se rechaza la hipótesis nula.

Todos los *gl* igual que en el capítulo 11; CM_{entre}, CM_{dentro}, y $F = S^2{}_{entre}$, $S^2{}_{dentro}$, y F del capítulo 11.

2. F necesario ($gl = 3, 5; p < 0{,}01$) = 12,06.

Fuente	SC	gl	CM	F
Intergrupal	298,89	3	99,63	41,51
Intragrupal	12	5	2,4	

Conclusión: Se rechaza la hipótesis nula.

3. i) a) $M_1 = 4$; $M_2 = 1$; $M_3 = 2$.
b) F necesario ($gl = 2, 6; p < 0{,}05$) = 5,14. (Nota: $GM = 2{,}33$).

Fuente	SC	gl	CM	F
Intergrupal	14	2	7	7,00
Intragrupal	6	6	1	

Conclusión: se rechaza la hipótesis nula.

c) $R^2 = 14/20 = 0{,}70$.

ii) a) $M^1 = 4$; $M^2 = 1$; $M^3 = 2$.
b) F necesario ($gl = 2, 6; p < 0{,}05$) = 5,14.
(Nota: $GM = 1{,}89$).

Fuente	SC	gl	CM	F
Intergrupal	12,89	2	6,45	4,85
Intragrupal	8,00	6	0,67	

Conclusión; no se rechaza la hipótesis nula.

c) $R^2 = 12{,}89/20{,}89 = 0{,}62$.

4. a) F necesario ($gl = 2, 9; p < 0{,}05$) = 4,26.

Fuente	SC	gl	CM	F
Intergrupal	84	2	42	9,95
Intragrupal	38	9	4,22	
Total		122	11	

Conclusión: se rechaza la hipótesis nula. Existe una diferencia significativa de autoestima entre los distintos tipos de maestros.

b) Tamaño del efecto (R^2) = 84/122 = 0,69.
c) Explicación: la lógica general es analizar si la variación de la autoestima entre las tres muestras podrían haber ocurrido más del 5% de las veces si, de hecho, las tres muestras hubieran sido tomadas al azar de tres poblaciones de maestros con la misma media del nivel de autoestima. En primer lugar, el procedimiento tiene en cuenta que si fueran precisamente tres muestras tomadas al azar de poblaciones con la misma media, la variación en cuanto al nivel de autoestima de cada grupo de maestros sería una base razonable para estimar la variación de la población. De modo similar, bajo estas condiciones, la variación de las medias de los grupos también sería una base para estimar la varianza general de población (esto se debe a que cualquier variación entre esas medias sólo puede ser el resultado de la variación entre los valores dentro de las tres poblaciones). Si ambas estimaciones son iguales, su razón debería ser 1 : 1, ó 1.

Pero supongamos que en realidad los grupos pertenecen a poblaciones con diferentes medias. En ese caso, la estimación de la variación a partir de las medias de los grupos debería ser mayor que aquella basada en la variación interna de cada grupo de maestros. Por lo tanto, la razón (si la variación intergrupal se ubica arriba) sería mayor a 1.

Dado que la cantidad de casos en cada grupo no es la misma, no es sencillo combinar la información de las tres muestras (o incluso determinar con precisión la variación entre los tres grupos) porque la información proporcionada por los grupos tiene diferente ponderación. Sin embargo, existe un procedimiento para simplificar este proceso. El procedimiento utiliza el principio que establece que, para cada observación, su desviación con respecto a la media general de todas las observaciones es igual a su desviación con respecto a la media de su propio grupo más la desviación de la media de su propio grupo con respecto a la media general. También resulta que (y puede ser probado matemáticamente), si elevamos al cuadrado cada una de estas diferentes desviaciones, la suma de todas las desviaciones cuadráticas con respecto a la gran media es igual a la suma de las desviaciones cuadráticas de cada registro con respecto a su media, más la suma de las desviaciones cuadráticas de la media del grupo de cada observación con respecto a la gran media. Al dividir las últimas dos sumas de cuadrados por los grados de libertad involucrados en cada cálculo, obtenemos las dos estimaciones de la varianza poblacional.

En el caso que estamos analizando, la suma de las desviaciones cuadráticas de la media del grupo de cada desviación, con respecto a la media general (de 6), era 84. Los grados de libertad son 2 porque sólo están involucradas las medias de tres grupos, y la estimación de varianza poblacional es 84/2, ó 42. De manera similar, la suma de las desviaciones cuadráticas de cada observación con respecto a la media de su grupo era 38. Los grados de libertad totales (la cantidad de registros de cada grupo menos 1, teniendo en cuenta todos los grupos) eran 9, y la estimación de varianza poblacional utilizando las variaciones dentro de cada grupo es de 4,22 (38/9).

La razón general entre la varianza poblacional, estimada sobre la base de la variación entre los grupos, y la varianza poblacional estimada sobre la base de la variación dentro de los grupos es 9,95 (42/4,22). Esa razón general se denomina razón *F*. Sucede que se conoce la distribución de todas las razones *F* posibles, aunque ésta varía según los grados de libertad en los que se basan las estimaciones de varianza intragrupal e intergupal. En este caso, buscando el punto de corte en el que un *F* ocurriría el 5% de las veces o menos, hayamos que el mismo es de 4,26 en una distribución de razones *F* basada en 2 y 9 grados de libertad. Ya que 9,95 es una razón *F* considerablemente mayor que el mínimo necesario de 4,26, podemos concluir que existe menos de un 5% de probabilidad de obtener esta variación entre nuestros grupos si los niveles de autoestima hubieran sido realmente tomados al azar de tres poblaciones de maestros con la misma media. Finalmente, se calcula una estimación del "tamaño del efecto", una indicación estandarizada de la cantidad de variación entre las medias. El procedimiento que se utilizó fue tomar las desviaciones cuadráticas totales de todos los registros con respecto a la gran media (que era 122) como una especie de línea de base de la variación a ser explicada. Después se calculó el porcentaje de ese total, que estaba explicado por las desviaciones cuadráticas de las medias de los grupos de las desviaciones, con respecto a la gran media (que era 84 y que es una indicación de la cantidad de variación entre grupos); el resultado fue 84/122, ó 69%. Es decir, el 69% de la variación de las desviaciones cuadráticas de las observaciones con respecto a la gran media está justificado por las desviaciones cuadráticas de las medias de sus grupos con respecto a la gran media, un tamaño del efecto bastante importante.

5. *F* necesario ($gl = 2, 7; p < 0{,}05$) = 4,74.

Fuente	*SC*	*gl*	*CM*	*F*
Intergrupal	66	2	33	9,62
Intragrupal	24	7	3,43	

Conclusión: se rechaza la hipótesis nula.
Explicación: *Véase* la respuesta al ejercicio 4 y las respuestas a ejercicios de capítulos anteriores.

6. La primera oración del resumen ofrece dos datos estadísticos clave con respecto a cada grupo estudiado, *M* y *SD*. *M* se refiere a la **media**, el promedio común de la cantidad de hermanos que tiene cada grupo. *SD* se refiere al ***desvío estándar*** de cada grupo, una indicación de la cantidad de variación ampliamente utilizada. En un sentido amplio, el desvío estándar es la variación promedio de la cantidad de hermanos con respecto al promedio del grupo. (En un sentido estricto, es la raíz cuadrada del promedio de las diferencias cuadráticas de la cantidad de hermanos de cada persona con respecto al promedio de su grupo).

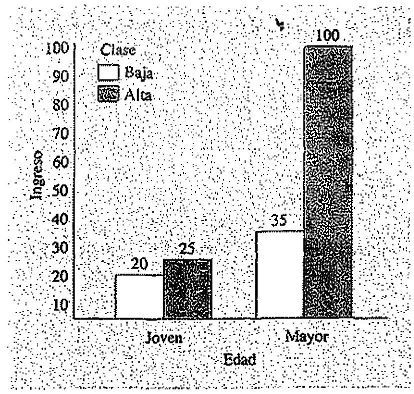
La línea citada justo antes del resumen ("*F*(2, 55)...") describe los resultados de una comparación estadística general entre medias de los tres grupos diferentes . Es decir, una cuestión de interés fundamental para estos investigadores es saber si las diferencias entre las medias es mayor de lo que se esperaría por casualidad, lo que se denomina "significación estadística" de la diferencia. (En este punto, explicaríamos la lógica del análisis de varianza y de la razón *F* siguiendo el estilo de la respuesta al ejercicio 3 del capítulo 11).

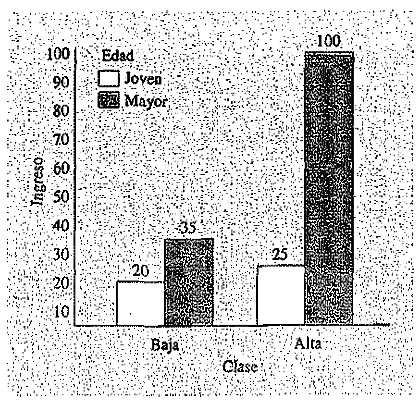
El resumen también se refiere a "comparaciones planificadas". Se trata de pruebas de significación de pares particulares de medias determinadas de antemano y basadas en la teoría. En este ejemplo, los investigadores planificaron una comparación entre los pro-sociales y los otros dos grupos juntos, y otra comparación entre los otros dos grupos. La primera comparación fue significativa (lo cual quiere decir que es altamente improbable que el estudio hubiera descubierto esta gran diferencia si, en efecto, las poblaciones involucaradas no fueran diferentes; de hecho habría menos de 5 posibilidades en mil). Sin embargo, la diferencia entre los individualistas y los competitivos no fue significativa, es decir que no es tan improbable que uno pudiera obtener una diferencia de ese tamaño si, de hecho, las poblaciones involucradas fueran idénticas).

Observando las cantidades específicas de hermanos involucrados, la conclusión es que el estudio sugiere que, en general (entre las personas como las estudiadas), las personas con una orientación pro-social probablemente tengan más hermanos que la gente que no tiene una orientación pro-social. Sin embargo, el estudio no es concluyente en cuanto a si entre las personas en general existe alguna diferencia entre la cantidad de hermanos –de los individualistas y de los competitivos.

Capítulo 13

1. i) a)

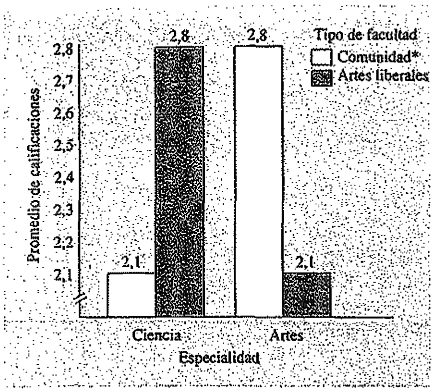


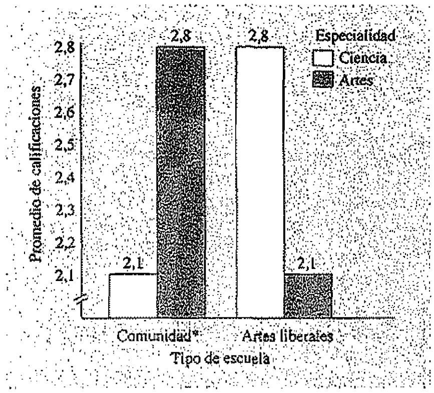


b) Efectos principales de la clase y de la edad; efecto interactivo.

c) El nivel de ingresos en general es mayor en la clase alta y en los individuos de mayor edad, pero la combinación de mayor edad y clase alta presenta un nivel de ingreso mayor de lo que se esperaría sólo por el efecto de alguna de las variables por separado.

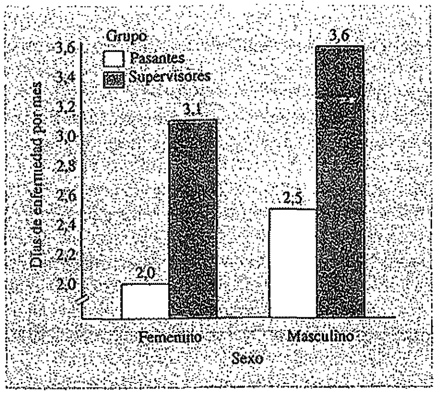
ii) a)

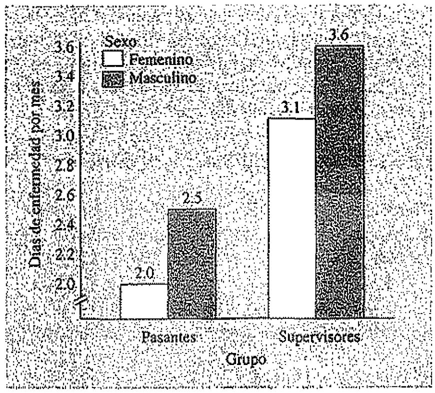




b) No hay efectos principales; efecto interactivo.
c) Ni el tipo de facultad ni el tipo de especialización, por sí solas, predicen calificaciones. Pero existe un patrón claro si uno analiza las combinaciones: las calificaciones correspondientes a alumnos especializados en arte de las facultades de la comunidad y a alumnos especializados en ciencia de las facultades de artes liberales, son más altas.

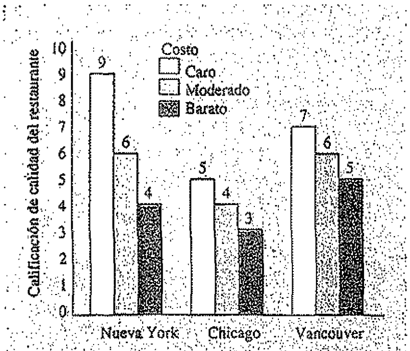
iii) a)

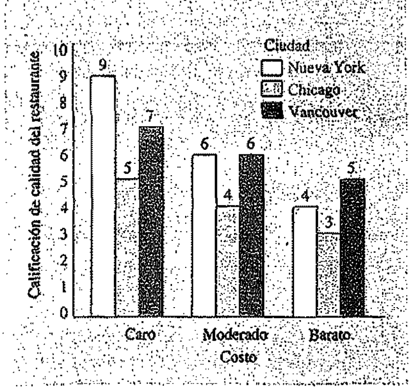




b) Ambos efectos principales son significativos; no hay interacción.
c) Las mujeres pierden menos días por mes que los hombres; los pasantes pierden menos días por mes que los supervisores. Cada combinación pierde la cantidad de días que esperaríamos conociendo el nivel en cada variable independiente por separado.

* N. de la trad.: *Community College*: Colegio que comprende dos años de universidad y es mantenido en parte por la comunidad a la cual sirve.





b) Efecto principal de la ciudad y el nivel de precio, más una interacción.

c) La calidad de los restaurantes es diferente en las distintas ciudades, siendo Nueva York la de más alta calidad y Chicago la de más baja calidad. La calidad de los restaurantes es diferente según los diferentes niveles de precio, siendo mejores los caros y peores los baratos. Sin embargo, los dos factores no se combinan simplemente, ya que el precio crea una mayor diferencia en Nueva York que en otras ciudades.

2. Las respuestas son sólo ejemplos.

Deporte

a)

Condición	*Béisbol*	*Fútbol americano*	*Baloncesto*	
Con programa de motivación	10	5	6	7
Sin programa de motivación	10	5	6	7
	10	5	6	

b)

Condición	*Béisbol*	*Fútbol americano*	*Baloncesto*	
Con programa de motivación	6	6	6	6
Sin programa de motivación	10	10	10	10
	8	8	8	

c)

Condición	*Béisbol*	*Fútbol americano*	*Baloncesto*	
Con programa de motivación	6	7	8	7
Sin programa de motivación	8	9	10	9
	7	8	9	

d)

Condición	*Béisbol*	*Fútbol americano*	*Baloncesto*	
Con programa de motivación	6	7	8	7
Sin programa de motivación	10	9	8	9
	8	8	8	

e)

Condición	*Béisbol*	*Fútbol americano*	*Baloncesto*	
Con programa de motivación	6	7	8	7
Sin programa de motivación	6	8	10	8
	6	7,5	9	

3. a) Efecto del deporte, 0,94; efecto de la condición, 0,97; efecto interactivo, 0,94.
   b) Deporte, *66*; *condición*, *54*; *interacción*, *66*. Por lo tanto, al menos 66 son necesarios.
4. a) Análisis de varianza:
   Punto de corte F para el efecto principal del diagnóstico ($gl = 1, 6; p < 0{,}05$) = 5,99.
   Punto de corte F para el efecto principal de la terapia ($gl = 2, 6; p < 0{,}05$) = 5,14.
   Punto de corte F para el efecto interactivo ($gl = 2, 6; p < 0{,}05$) = 5,14.

Terapia A

		X	$(X - GM)^2$	$(X - M)^2$	$(M_{Columna} - GM)^2$	$(M_{fila} - GM)^2$	Int^2
I		6	0	4	1	9	0
		2	16	4	1	9	0
	M	4	16	8	2	18	0
II		11	25	1	1	9	0
		9	9	1	1	9	0
	M	10	34	2	2	18	0

$M_{Columna}$ 7

Terapia B

		X	$(X - GM)^2$	$(X - M)^2$	$(M_{Columna} - GM)^2$	$(M_{fila} - GM)^2$	Int^2
I		3	9	1	1	9	0
		1	25	1	1	9	0
	M	2	34	2	2	18	0
II		7	1	1	1	9	0
		9	9	1	1	9	0
	M	8	10	2	2	18	0

$M_{Columna}$ 5

Terapia C

		X	$(X - GM)^2$	$(X - M)^2$	$(M_{Columna} - GM)^2$	$(M_{fila} - GM)^2$	Int^2
I		2	16	1	0	9	0
		4	4	1	0	9	0
	M	3	20	2	0	18	0
II		8	4	1	0	9	0
		10	16	1	0	9	0
	M	9	20	2	0	18	0

$M_{Columna}$ 6

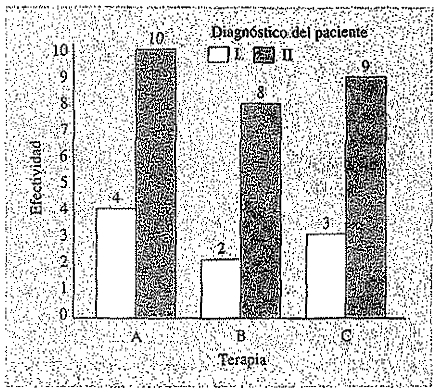
$SS_{total} = 16 + 34 + 20 + 34 + 10 + 20 = 134.$
$SS_{dentro} = 8 + 2 + 2 + 2 + 2 + 2 = 18.$
$SS_{columnas} = 2 + 2 + 0 + 2 + 2 + 0 = 8.$
$SS_{filas} = 18 + 18 + 18 + 18 + 18 + 18 = 108.$
$SS_{interacción} = 0 + 0 + 0 + 0 + 0 + 0 = 0.$

Fuente	*SC*	*gl*	*CM*	*F*	
Terapia	8	2	4	1,33	No se rechaza la hipótesis nula
Diagnóstico	108	1	108	36	Se rechaza la hipótesis nula
Interacción	0	2	0	0	No se rechaza la hipótesis nula
Interior de casillas	18	6	3		

b) Tabla de medias de casilla y marginales

Medias:

	A	B	C	
I	4	2	3	3
II	10	8	9	9
	7	5	6	



c) Tamaños de efecto:
$R^2_{Columnas} = 8/(134 - 108 - 0) = 8/26 = 0{,}31$
$R^2_{fila} = 108/(134 - 8 - 0) = 108/126 = 0{,}86$
$R^2_{Interacción} = 0/(134 - 8 - 108) = 0/18 = 0$

d) Explicación: los resultados indican que existe una diferencia *significativa en la efectividad* entre las dos categorías de diagnóstico; la terapia es más efectiva para aquellos con diagnóstico II. Sin embargo, no existe diferencia significativa entre los tipos de terapia, y los tipos de terapia no presentan una diferencia de efectividad significativa en los distintos tipos de diagnóstico. El tamaño del efecto significativo es extremadamente grande.

5. a) Análisis de varianza:
Punto F de corte necesario para el efecto principal de la simpatía ($gl = 1, 8; p < 0{,}05$) = 5,14.
Punto F de corte necesario para el efecto principal del nerviosismo ($gl = 1{,}8$; $p < 0{,}05$) = 5,14.
Punto F de corte necesario del efecto interactivo ($gl = 1, 8; p < 0{,}05$) = 5,14.

Simpatía

	X	$(X - GM)^2$	$(X - M)^2$	$(M_{Fila} - GM)^2$	$(M_{Columna} - GM)^2$	Int^2
Nerviosismo	7	4	0	0	0	4
	8	9	1	0	0	4
	6	1	1	0	0	4
M	7	14	2	0	0	12
Ausencia de Nerviosismo	3	4	0	0	0	4
	3	4	0	0	0	4
	3	4	0	0	0	4
	3	12	0	0	0	12
$M_{Columna}$	5					

Ausencia de simpatía

	X	$(X - GM)^2$	$(X - M)^2$	$(M_{Fila} - GM)^2$	$(M_{Columna} - GM)^2$	Int^2	
Nerviosismo	3	4	0	0	0	4	
	4	1	1	0	0	4	
	2	9	1	0	0	4	
M	3	14	2	0	0	12	5
Ausencia de Nerviosismo	7	4	0	0	0	4	
	5	0	4	0	0	4	
	9	16	4	0	0	4	
	7	20	8	0	0	12	5
$M_{Columna}$	5				GM = 5		

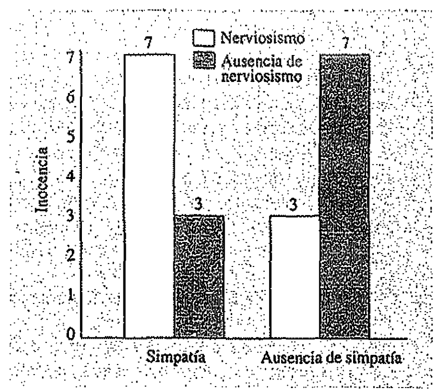
$SC_{Total} = 60.$
$SC_{Dentro} = 12.$
$SC_{Columnas} = 0.$
$SC_{Fila} = 0.$
$SC_{Interacción} = 48.$

Fuente	SC	gl	CM	F	
Simpatía	0	1	0	0	No se rechaza la hipótesis nula.
Nerviosismo	0	1	0	0	No se rechaza la hipótesis nula.
Interacción	48	1	48	24	Se rechaza la hipótesis nula.
Dentro de casillas	12	8	2		

b) Tabla de medias de casilla y marginales, y gráfico:

Medias:

	Simpatía	Ausencia de simpatía	
Nervioso	7	3	5
Ausencia de Nerviosismo	3	7	5
	5	5	



c) Tamaño del efecto:
$R^2_{Columnas} = 0 / (60 - 0 - 48) = 0/12 = 0$
$R^2_{fila} = 0 / (60 - 0 - 48) = 0/12 = 0$
$R^2_{Interacción} = 48 / (60 - 0 - 0) = 48/60 = 0{,}80$

d) Explicación: los resultados indican que existe una interacción significativa entre el nerviosismo y la simpatía: cuando el acusado es simpático, tiene más probabilidades de ser calificado inocente si está nervioso; pero si no es simpático , tiene más probabilidades de ser calificado inocente si no está nervioso. (Tal vez uno puede sentir empatía con el nerviosismo de una persona simpática en el estrado, y si la persona no estuviera nerviosa, uno podría sospechar algo raro. En una persona que no es simpática, el nerviosismo puede ser una advertencia de que es culpable, pero si no está nervioso, sugiere que no tiene nada que ocultar). No hubo efecto significativo general en cuanto a la simpatía o a la falta de ella, o en cuanto al nerviosismo o la falta del mismo, aunque debido al pequeño tamaño de las muestras utilizadas, el no poder rechazar la hipótesis nula no debería tomarse como prueba de que no existe tal efecto.

El cálculo de la significación en este experimento es muy parecido al análisis de varianza de un criterio utilizando el método del modelo estructural. Los grados de libertad y la suma de cuadrados intragrupales se calculan de la forma acostumbrada, considerando a cada una de las cuatro casillas como su propio grupo. Sin embargo, en este caso, el desvío intergrupal se divide en partes. Una parte tiene en cuenta la variación intergrupal de la simpatía y la ausencia de la misma. Los desvíos se calculan para cada participante tomando la media de todos los participantes en esa condición de simpatía y ausencia de simpatía a la que pertenece el participante, y restándole la gran media. Luego, los desvíos se elevan al cuadrado y se suman para obtener la suma de cuadrados. Después se repite el proceso con la condición de nerviosismo y ausencia de nerviosismo. Los grados de libertad para cada condición son la cantidad de niveles menos uno. Por ejemplo, dado que hay dos niveles de simpatía (simpatía y ausencia de simpatía), esta condición tiene 1 grado de libertad.

Aún queda una parte correspondiente al efecto intergrupal que tiene en cuenta las variaciones intergrupales de las medias de cada uno de los cuatro subgrupos, que no son simplemente el resultado de sumar los efectos de la simpatía y el nerviosismo. Es decir, toda variación entre los grupos de simpatía, que difiere según el grupo de nerviosismo al que pertenezcan. El desvío para este efecto interactivo se encuentra tomando el desvío de cada registro con respecto a la gran media general, y restándole los otros tres desvíos (el del registro menos la media de su grupo, y los de la media de simpatía menos la gran media, y la media de nerviosismo menos la gran media). Después se elevan al cuadrado esos desvíos restantes y se suman para convertirse en la suma de cuadrados de la interacción. Los grados de libertad son los que restan del total de grados de libertad intergrupales. Como hay cuatro subgrupos, los *gl* intergrupales = 3, y dado que hemos utilizado 1 para simpatía o ausencia de simpatía y 1 para nerviosismo o ausencia de nerviosismo, queda 1 *gl* para la interacción.

6. En este estudio hubo dos hallazgos importantes. Primero, como se esperaba, los participantes con estereotipos extremos en cuanto a que los agentes RRPP son extrovertidos, comparados con los participantes con estereotipos moderados, describieron a los agentes RRPP como más extrovertidos. Este resultado fue estadísticamente significativo; por consiguiente, podemos confiar en que el patrón del resultado se aplica no sólo a las personas estudiadas en particular sino a las personas en general que sean similares a las estudiadas. (Más precisamente, hemos calculado que si no existiera diferencia promedio en la población en general, entre personas con estereotipos extremos y moderados existiría menos de un 0,0001 de probabilidad de que este experimento produjera un resultado tan fuerte como el obtenido). Más aún, con relación al tamaño de efecto (proporción de varianza explicada) típicamente encontrado en los estudios psicológicos, la diferencia obtenida fue considerable. (Utilizando la fórmula basada en los F, $R^2 = (38{,}94)(1) / [(38{,}94)(1) + 42] = 0{,}48$).

   En segundo lugar, y de suma importancia, es que sorprendentemente la tendencia fue mucho más fuerte en los participantes a quienes se les dio una descripción de un agente RRPP en particular que era altamente introvertido. El resultado también fue estadísticamente significativo. (En este caso, la posibilidad de obtener un resultado tan fuerte, si en la población en general no hubiera una tendencia promedio del tipo observado, era menor al 5%). El patrón de este resultado también tenía un tamaño del efecto bastante grande con relación a lo que usualmente se ve en los estudios psicológicos ($R^2 = (5{,}69)(1) / [(5{,}69)(1) + 42] = 0{,}12$).

   En líneas generales, las personas expuestas al extremo introvertido tendieron a dar mayores calificaciones de extroversión. El resultado tuvo significación estadística "marginal", es decir que se encontraba en el límite de ser demasiado improbable que sucediera si no existiera verdadera diferencia promedio en la población. Más aún, el resultado no es muy interesante, ya que, como se puede observar en el gráfico, se debe enteramente a los participantes con estereotipos extremos, y si algún efecto se observa en los participantes de estereotipos moderados es en realidad un patrón de efecto contrario.

Capítulo 14

1. a) Punto de corte χ^2 necesario
 ($gl = 5 - 1 = 4$, 5%) = 9,488.

Categoría	O	Esperado	$O-E$	$(O-E)^2$	$(O-E)^2/E$
A	19	(0,2)(50) = 10	9	81	8,10
B	11	(0,2)(50) = 10	1	1	0,10
C	10	(0,4)(50) = 20	−10	100	5,00
D	5	(0,1)(50) = 5	0	0	0.00
E	5	(0,1)(50) = 5	0	0	0.00
Total	50	(0,1)(50) = 50	0		$\chi^2 = 13,20$

Conclusión: se rechaza la hipótesis nula.

b) Punto de corte χ^2 necesario
 ($gl = 3 - 1 = 2$, 5%) = 5,992.

Categoría	O	Esperado	$O-E$	$(O-E)^2$	$(O-E)^2/E$
I	100	(0,3)(300) = 90	10	100	1,11
II	100	(0,5)(300) = 150	-50	2,500	16,67
III	100	(0,2)(300) = 60	40	1,600	26,67
Total	300	300	0		$x^2 = 44,45$

Conclusión: se rechaza la hipótesis nula.

c) Punto de corte χ^2 necesario
 ($gl = 4 - 1 = 3$, 5%) = 7,815.

Categoría	O	Esperado	$O-E$	$(O-E)^2$	$(O-E)^2/E$
1	38	(100/500)(200) = 40	-2	4	0,10
2	124	(300/500)(200) = 120	4	16	0,13
3	22	(50/500)(200) = 20	2	4	0,20
4	16	(50/500)(200) = 20	-4	16	0,80
Total	200	200	0		$\chi^2 = 1,23$

Conclusión: no se rechaza la hipótesis nula.

d) Punto de corte χ^2 necesario
 ($gl = 3 - 1 = 2$, 5%) = 5,992.

Categoría	O	Esperado	$O-E$	$(O-E)^2$	$(O-E)^2/E$
Artes	37	30	7	49	1,63
Ciencias	21	30	-9	81	2,70
Humanidades	32	30	2	4	0,13
Total	90	900			$\chi^2 = 4,46$

Conclusión: no se rechaza la hipótesis nula.

2. a) Pasos de la prueba de hipótesis:
 1. Replantear el problema en función de hipótesis de investigación e hipótesis nula de las poblaciones.
 Población 1: pacientes como los de la clínica de psicoterapia del ejemplo.
 Población 2: pacientes para quienes las distintas temporadas no implican ninguna diferencia con respecto al momento en que comienzan la psicoterapia.

La hipótesis de investigación establece que la distribución entre las distintas temporadas, con respecto al momento en que los pacientes comienzan la psicoterapia, es diferente entre las dos poblaciones. La hipótesis nula establece que la distribución entre las distintas temporadas, con respecto al momento en que los pacientes comienzan la psicoterapia, no es diferente entre las dos poblaciones.

2. Determinar las características de la distribución comparativa.
Distribución de chi-cuadrados con 3 grados de libertad ($gl = 4 - 1 = 3$).

3. Determinar el punto de corte en la distribución comparativa, a partir del cual se debería rechazar la hipótesis nula.
Nivel 0,05, $gl = 3$: $\chi^2 = 7,815$.

4. Determinar el valor muestral en la distribución comparativa.

Temporada	O	Esperado	$O-E$	$(O-E)^2$	$(O-E)^2/E$
Invierno	28	(1/4)(128) = 32	−4	16	0,50
Primavera	33	(1/4)(128) = 32	1	1	0,03
Verano	16	(1/4)(128) = 32	−16	256	8,00
Otoño	51	(1/4)(128) = 32	19	361	11,28
Total	128	128	0		$\chi^2 = 19,81$

5. Comparar los valores obtenidos en los pasos 3 y 4 para decidir si se rechaza o no la hipótesis nula.
χ^2 en el paso 4 (19,81) es mayor que el punto de corte del paso 3 (7,815). Por lo tanto, se rechaza la hipótesis nula; se sostiene la hipótesis de investigación.

b) Explicación: si las temporadas no marcaran una diferencia esperaríamos aproximadamente un 25% de nuevos pacientes cada temporada (basándonos en el año anterior, el 25% del total de 128 es igual a 32). ¿Las cantidades reales de cada temporada del año anterior son tan diferentes a estas expectativas que deberíamos concluir que, en general, las cantidades de nuevos pacientes no se distribuyen en forma pareja entre las temporadas?
 El chi-cuadrado es un indicador del grado de discrepancia entre resultados observados y esperados. Para cada categoría (las cuatro estaciones en este caso), calculamos la diferencia, la elevamos al cuadrado y la dividimos por la cantidad esperada; después sumamos los resultados. En el invierno, 28 menos 32 es −4;

elevado al cuadrado es 16; dividido 32 es 5. Si hacemos lo mismo para las otras tres estaciones y sumamos los cuatro resultados obtenemos un chi-cuadrado total de 19,81. (El chi-cuadrado utiliza diferencias cuadráticas para que el resultado no se vea afectado por la dirección de las diferencias. Se divide por la cantidad esperada para reducir el impacto en el resultado de la cantidad ordinaria de casos).

Los estadísticos han determinado matemáticamente qué sucedería si tomamos una cantidad infinita de muestras de la población, con una proporción fija de casos en cada categoría, y calculamos el chi-cuadrado para cada una de esas muestras. La distribución de esos chi-cuadrados depende sólo de la cantidad de categorías libres para incluir diferentes valores esperados. (Ya que la cantidad total esperada es la cantidad total de casos, si conocemos la cantidad esperada para tres categorías cualesquiera, la cantidad esperada para la cuarta es fácil de determinar. Una tabla de la distribución chi-cuadrado, con tres categorías libres de variar, muestra que existe sólo un 5% de posibilidades de obtener un chi-cuadrado de 7,815 ó mayor. Como nuestro chi-cuadrado es mayor que ese número, el resultado observado difiere del esperado más de lo que razonablemente esperaríamos que lo hiciera por casualidad; la cantidad de nuevos pacientes, a la larga, probablemente no sea la misma durante las cuatro estaciones.

3. a) Cálculo

Categoría	O	Esperado	O–E	(O–E)²	(O–E)²/E
Cambio Promedio	43	(1/2) (57) = 28,5	14,5	210,25	7,38
Cambio No Percibido	14	(1/2) (57) = 28,5	-14,5	210,25	7,38
				$\chi^2 =$	14,76

b) Explicación: similar a la respuesta al ejercicio 2b.

4. a) $gl = (N_{\text{Columnas}} - 1)(N_{\text{filas}} - 1) = (2-1)(2-1) = 1$; Punto de corte χ^2 necesario ($gl = 1$, 1%) = 6,635.

10 (13)	16 (13)	26 (50%)
16 (13)	10 (13)	26 (50%)
26	26	52

$$\chi^2 = \frac{(10-13)^2}{13} + \frac{(16-13)^2}{13} + \frac{(16-13)^2}{13} + \frac{(10-13)^2}{13}$$

$$= 0,69 + 0,69 + 0,69 + 0,69 = 2,76.$$

No se rechaza la hipótesis nula.

$$\phi = \sqrt{\chi^2/N} = \sqrt{2,6/52} = \sqrt{0,53} = 0,23.$$

b) $gl = (N_{\text{Columnas}} - 1)(N_{\text{fila}} - 1) = (2-1)(2-1) = 1$; Punto de corte X^2 necesario (gl = 1, 1%) = 6,635.

100 (103)	106 (103)	206 (50%)
106 (103)	100 (103)	206 (50%)
206	206	412

$$\chi^2 = \frac{(100-103)^2}{103} + \frac{(106-103)^2}{103} + \frac{(106-103)^2}{103} + \frac{(100-103)^2}{103}$$

$$= 0,09 + 0,09 + 0,09 + 0,09 = 0,36.$$

No se rechaza la hipótesis nula.

$$\phi = \sqrt{0,36/412} = \sqrt{0,0009} = 0,03.$$

c) $gl = (N_{\text{Columnas}} - 1)(N_{\text{filas}} - 1) = (2-1)(2-1) = 1$; Punto de corte χ^2 necesario ($gl = 1$, 1%) = 6,635.

$$\chi^2 = \frac{(100-130)^2}{130} + \frac{(160-130)^2}{130} + \frac{(160-130)^2}{130} + \frac{(100-130)^2}{130}$$

$$= 6,92 + 6,92 + 6,92 + 6,92 = 27,68.$$

Se rechaza la hipótesis nula.

$$\phi = \sqrt{27,68/520} = \sqrt{0,0532} = 0,23.$$

d) $gl = (N_{\text{Columnas}} - 1)(N_{\text{filas}} - 1) = (3-1)(2-1) = 2$; Punto de corte χ^2 necesario ($gl = 2$, 1%) = 9,211.

10 (13)	16 (13)	10 (10)	36 (50%)
16 (13)	10 (13)	10 (10)	36 (50%)
26	26	20	72

$$\chi^2 = \frac{(10-13)^2}{13} + \frac{(16-13)^2}{13} + \frac{(16-13)^2}{13} + \frac{(10-13)^2}{13} + \frac{(10-10)^2}{10} + \frac{(10-10)^2}{10}$$

$$= 0,69 + 0,69 + 0,69 + 0,69 + 0 + 0 = 2,76.$$

No se rechaza la hipótesis nula.

$$\phi \text{ de Cramer} = \sqrt{\chi^2/(N)(gl_{\text{Menor}})} = \sqrt{2,76/(72)(1)} = \sqrt{0,0383} = 0,20.$$

e) $gl = (N_{\text{Columnas}} - 1)(N_{\text{filas}} - 1) = (3-1)(2-1) = 2$; Punto de corte χ^2 necesario ($gl = 2$, 1%) = 9,211.

10 (13)	16 (13)	16 (16)	42 (50%)
16 (13)	10 (13)	16 (16)	42 (50%)
26	26	32	84

$$\chi^2 = \frac{(10-13)^2}{13} + \frac{(16-13)^2}{13} + \frac{(16-13)^2}{13} + \frac{(10-13)^2}{13}$$

$$+ \frac{(16-16)^2}{16} + \frac{(16-16)^2}{16}$$

$$= 0{,}69 + 0{,}69 + 0{,}69 + 0{,}69 + 0 + 0 = 2{,}76.$$

No se rechaza la hipótesis.

ϕ de Cramer $= \sqrt{2{,}76/(84)(1)} = \sqrt{0{,}0329} = 0{,}18$.

f) $gl = (N_{\text{Columnas}} - 1)(N_{\text{filas}} - 1) = (3-1)(2-1) = 2$; χ^2 necesario $(gl = 2, 1\%) = 9{,}211$.

10 (12)	16 (12)	10 (12)	36 (46%)
16 (14)	10 (14)	16 (14)	42 (54%)
26	26	26	78

$$\chi^2 = \frac{(10-12)^2}{12} + \frac{(16-12)^2}{12} + \frac{(10-12)^2}{12} + \frac{(16-14)^2}{14}$$

$$+ \frac{(10-14)^2}{14} + \frac{(16-14)^2}{14}$$

$$= 0{,}33 + 1{,}33 + 0{,}33 + 0{,}29 + 1{,}14 + 0{,}29 = 3{,}71$$

No se rechaza la hipótesis nula.

ϕ de Cramer $= \sqrt{3{,}71/(78)(1)} = \sqrt{0{,}0476} = 0{,}22$.

5. a) Pasos de la prueba de hipótesis:

1. Replantear el problema en función de hipótesis de investigación e hipótesis nula de las poblaciones.

Población 1: alumnos como los entrevistados.

Población 2: alumnos para quienes el tipo de artefacto utilizado en su hogar es independiente de la utilización de lapicera o lápiz cuando toman apuntes en clase.

La hipótesis de investigación establece que las dos poblaciones son diferentes (el tipo de artefacto utilizado en los hogares no es independiente del hecho de utilizar lapicera o lápiz al tomar apuntes en clase). La hipótesis nula establece que las dos poblaciones son iguales (el tipo de artefacto utilizado en los hogares es independiente del hecho de utilizar lapicera o lápiz al tomar apuntes en clase).

2. Determinar las características de la distribución comparativa.

Distribución chi-cuadrado con dos grados de libertad.

$gl = (N_{\text{Columnas}} - 1)(N_{\text{filas}} - 1) = (3-1)(2-1) = 2$

3. Determinar el punto de corte en la comparación

Nivel 0,05, $gl = 2$: $\chi^2 = 5{,}992$.

4. Determinar el registro muestral en la distribución comparativa.

Artefacto utilizado en los hogares

Elemento utilizado para tomar apuntes	*Máquina de escribir*	*Procesador de textos*	*Ninguno*	
Lapicera	42 (39)	62 (65)	26 (26)	130 (65%)
Lápiz	18 (21)	38 (35)	14 (14)	70 (35%)
	60	100	40	200

$$\chi^2 = \frac{(42-39)^2}{39} + \frac{(62-65)^2}{65} + \frac{(26-26)^2}{26} + \frac{(18-21)^2}{21}$$

$$+ \frac{(38-35)^2}{35} + \frac{(14-14)^2}{14}$$

$$= 0{,}23 + 0{,}14 + 0 + 0{,}43 + 0{,}26 = 1{,}06$$

5. Comparar los valores obtenidos en los pasos 3 y 4 para decidir si se rechaza o no la hipótesis nula.

χ^2 en el paso 4 (1,06) es menos extremo que el punto de corte del paso 3 (5,992).

Por lo tanto, no se rechaza la hipótesis nula; el estudio no es concluyente.

b) ϕ de Cramer's $= \sqrt{[1{,}06/(200)(1)]} = \sqrt{0{,}0053} = 0{,}07$.

c) Explicación: en este ejemplo, el 65% de todos los sujetos utilizan lapiceras para tomar apuntes. Por lo tanto, si el hecho de utilizar lapicera o lápiz no está relacionado con el artefacto utilizado en los hogares, el 65% de las personas que forman cada categoría referida al artefacto que se utiliza en el hogar utilizaría lapiceras para tomar apuntes. Por ejemplo, esperaríamos que 39 de los 60 alumnos que utilizan máquina de escribir utilizarán lapicera al tomar apuntes. ¿Los resultados de la encuesta, son lo suficientemente diferentes de estas expectativas como para que concluyéramos que lo que los alumnos utilizan para tomar apuntes está relacionado con el artefacto que emplean para escribir cuando están en sus hogares?

El chi-caudrado es una medida del grado de discrepancia entre los resultados observados y esperados. Calculamos la diferencia entre lo observado y lo esperado en cada combinación de la estructura 2 x 3, elevamos esa diferencia al cuadrado y la dividimos por la cantidad esperada; luego sumamos los resultados. En la combinación lapicera-máquina de escribir, 42 menos 39 es 3, elevado al cuadrado es 9, dividido 39 es 0,23. Al realizar el mismo proceso para las otras cinco combinaciones y sumarlas obtenemos 1,06. (Los chi-cuadrados utilizan diferencias elevadas al cuadrado para que el resultado no se vea afectado por las direcciones de las diferencias. Además, la diferencia cuadrática se divide por la cantidad esperada para adaptar el impacto de las cantidades relativamente diferentes esperadas para cada combinación).

Los estadísticos han determinado matemáticamente lo que sucedería si tomáramos una cantidad infinita de muestras de una población, con una proporción fija de personas en cada una de las distintas agrupaciones, y calculáramos el chi-cuadrado de cada una de esas muestras. La distribución de esos chi-cuadrados depende sólo de la cantidad de agrupaciones libres para adoptar diferentes valores esperados. (Si de cada una de las categorías en las que se divide la variable "artefactos utilizados en el hogar" conocemos la cantidad de alumnos que toman apuntes con lapicera, es fácil determinar la cantidad de alumnos que toma apuntes con lápiz. Y si conocemos dos de las tres categorías de la variable "artefactos utilizados en los hogares", correspondientes al grupo que utiliza lapicera, la tercera categoría es fácil de determinar porque debe sumar el total de alumnos que utilizan lapicera. Por lo tanto, sólo dos combinaciones son "libres de variar").

Una tabla de la distribución de chi-cuadrados, para el caso en que dos agrupaciones son libres de variar, muestra que existe sólo un 5% de posibilidades de obtener un chi-cuadrado de 5,992 ó mayor. Debido a que nuestro chi-cuadrado es menor a ese número, las cantidades observadas en cada categoría difieren de las cantidades esperadas menores a lo necesario para poder rechazar la idea de que el elemento que las personas utilizan para tomar apuntes no está relacionado con el artefacto que utilizan para escribir cuando están en sus hogares. La encuesta no es concluyente.

Sin embargo, podemos estimar el grado real de relación, dentro de este grupo, entre el elemento utilizado en clase y el artefacto utilizado en los hogares. El procedimiento mencionado se denomina "phi de Cramer", y se calcula dividiendo el chi-cuadrado calculado por la cantidad de personas que incluye el análisis, sacando luego la raíz cuadrada del resultado. En el ejemplo que analizamos el resultado es 0,07.

El estadístico que mencionamos en el párrafo anterior se extiende del 0 (ausencia de relación) al 1 (relación perfecta, conocer la situación de una persona en una de las dimensiones, como por ejemplo, saber qué utiliza para escribir en clase permitiría predecir perfectamente su situación en la otra dimensión, tal como el elemento que utiliza para escribir en su hogar). Por lo tanto, 0,07 es un número bastante bajo. (De hecho, el phi de Cramer es comparable con lo que se denomina un coeficiente de correlación, y en psicología 0,07 es un valor muy bajo con respecto a las correlaciones encontradas en la mayoría de los estudios). Viéndolo de otro modo, podemos preguntar, si realmente existe una relación moderada, ¿cuáles son las posibilidades de que todo el proceso realizado diera como resultado una conclusión positiva? Los estadísticos han desarrollado tablas que nos indican esa probabilidad y, en este caso, habría un 97% de probabilidad. Por lo tanto, dado el resultado obtenido, si existe alguna relación, casi seguramente es bastante pequeña.

6. a) Los cinco pasos de la prueba de hipótesis deberían realizarse de forma similar a los indicados en la respuesta al ejercicio 5 que aparece anteriormente. Los cálculos y resultados clave se indican a continuación:

$gl = (N_{\text{Columnas}} - 1)(N_{\text{filas}} - 1) = (3 - 1)(3 - 1) = 4$;
Punto de corte χ^2 necesario ($gl = 4$, 5%) = 9,488.

Comunidad

	A	B	C	Total
A favor	12 (9,8)	6 (4,2)	3 (7)	21 (23,33%)
En contra	18 (16,8)	3 (7,2)	15 (12)	36 (40,00%)
No emite opinión	12 (15,4)	9 (6,6)	12 (11)	33 (36,67%)
Total	42	18	30	90

$$\chi^2 = \frac{(12-9,8)^2}{9,8} + \frac{(6-4,2)^2}{4,2} + \frac{(3-7)^2}{7} + \frac{(18-16,8)^2}{16,8}$$
$$+ \frac{(3-7,2)^2}{7,2} + \frac{(15-12)^2}{12} + \frac{(12-15,4)^2}{15,4} + \frac{(9-6,6)^2}{6,6}$$
$$+ \frac{(12-11)^2}{11}$$
$$= 0,49 + 0,77 + 2,29 + 0,09 + 2,45 + 0,75 + 0,75 + 0,87 + 0,09$$
$$= 8,55$$

No se rechaza la hipótesis nula.

b) f de Cramer = $\sqrt{8,55/(90)(2)}$ = $\sqrt{8,55/180}$ = $0,05 = 0,22$.

Potencia para un tamaño de efecto pequeño = 0,11; mediano = 0,66; grande = 0,99. (Sobre la base de N = 100).

c) Explicación: *Véase* la respuesta al ejercicio 5c.

7. a) Cálculo de χ^2

	Han tratado	No han tratado	Total
≤ 40 hrs.	37 (52,9)	51 (35,2)	88 (15,80%)
40–49 hrs.	70 (70,4)	47 (46,8)	117 (21,00%)
≥ 50 hrs.	228 (212,1)	125 (141,2)	353 (63,30%)
Total	335	223	558 (100,10%)

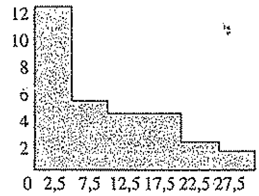
$$\chi^2 = \Sigma\, \frac{(O-E)^2}{E}$$
$$= \frac{(37-52,9)^2}{52,9} + \frac{(51-35,2)^2}{35,2} + \frac{(70-70,4)^2}{70,4}$$
$$+ \frac{(47-46,8)^2}{46,8} + \frac{(228-212,1)^2}{212,1} + \frac{(125-141,2)^2}{141,2}$$
$$= 4,78 + 7,09 + 0,00 + 0,00 + 1,19 + 1,86 = 14,92$$

b) ϕ de Cramer = $\sqrt{[14,92/(558)(1)]}$ = $\sqrt{0,027}$ = 0,16; tamaño del efecto pequeño.

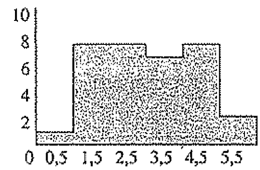
c) El principal hallazgo es que las proporciones de médicos que han tratado pacientes con HIV/SIDA es mucho menor dentro del grupo de médicos que ejercen menos de 40 horas por semana. De todos modos, se trata de un tamaño del efecto pequeño. La explicación para una persona que nunca ha tomado un curso de estadística sería similar a la respuesta al ejercicio 5c anterior.

Capítulo 15

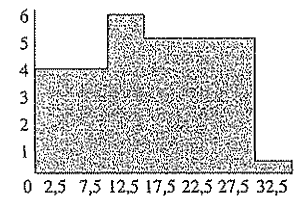
1.



(a)



(b)



(c)

Datos sin transformar		**Raíz cuadrada**		**Rangos**	
0–4	12	0–0,9	1	0–4,9	4
5–9	5	1–1,9	7	5–9,9	4
10–14	4	2–2,9	7	10–14,9	6
15–19	4	3–3,9	6	15–19,9	5
20–24	3	4–4,9	7	20–24,9	5
25–30	2	5–5,9	2	25–29,9	5
				30–34,9	1

Original	f	Rango	$\sqrt{\ }$	Original	f	Rango	$\sqrt{\ }$
0/	1	1	0,0	14			
1///	3	3	1,0	15			
2///	3	6	1,4	16//	2	22,5	4,0
3/	1	8	1,7	17/	1	24	4,1
4////	4	10,5	2,0	18/	1	25	4,2
5/	1	13	2,2	19			
6/	1	14	2,4	20/	1	26	4,5
7/	1	15	2,6	21/	1	27	4,6
8				22			
9//	2	16,5	3,0	23/	1	28	4,8
10//	2	18,5	3,2	24			
11/	1	20	3,3	25/	1	29	5,0
12/	1	21	3,5	26			
13				27			
14				28/	1	30	5,3
15							

2. Probablemente no normal: a) asimétrica hacia la derecha, b) bimodal, c) asimétrica hacia la derecha.
3. a) y b) Punto de corte t necesario (dos colas, $p < 0{,}05$, gl = 8) = 2,306

Valores a los que se le aplicó la transformación raíz cuadrada

	Grupo A	*Grupo B*
	1,1	1,4
	1,6	3,0
	2,1	2,4
	1,9	2,6
	2,7	2,2
$M =$	1,88	2,32
$S^2 =$	0,35	0,35
$S^2_{Combinada} =$	0,35	
$S^2_M =$	0,07	0,07

$S^2_{Diferencia} = 0{,}07 + 0{,}07 = 0{,}14$; $S_{Diferencia} = 0{,}37$
$t = (1{,}88 - 2{,}32)/0{,}37 = -1{,}19$
conclusión: no se rechaza la hipótesis nula.

c) Explicación: no habría sido adecuado realizar una prueba t con los números tal como estaban (sin transformarlos). Las distribuciones de las muestras eran tan asimétricas para ambos grupos idiomáticos que parecía probable que la distribución poblacional también fuera considerablemente asimétrica. En ese caso, no se cumpliría el supuesto para la prueba t que establece que las distribuciones poblacionales implícitas son normales. Por lo tanto, se calculó la raíz cuadrada de cada observación. A través de ese proceso se obtuvo la posibilidad de crear una distribución muestral mucho más cercana a lo normal, y que, por lo tanto, probablemente sugiere que la distribución poblacional de raíces cuadradas de los tamaños de las familias está distribuida de forma prácticamente normal. Somos conscientes de que calcular la raíz cuadrada de cada tamaño familiar distorsiona su significado directo. Pero el impacto causado a los individuos de la familia por cada hijo adicional probablemente no sea igual. Es decir, no tener ningún hijo y tener uno provoca un enorme impacto. Pasar de tener 1 a tener 2 provoca un impacto menor, y pasar de tener 7 a tener 8 probablemente provoca una diferencia mucho menor para la familia.

De todos modos, después de haber calculado la raíz cuadrada de cada observación, se realizó una prueba t común para medias independientes. El resultado no fue concluyente; no se pudo rechazar la hipótesis nula. (Y dado que el tamaño de la muestra era tan pequeño, la potencia probablemente también era baja, haciendo difícil deducir algún significado del hecho de no haber podido rechazar la hipótesis nula).

4. a) y b)

Observación:
201 523 614   136 340 301   838 911 1.007
Rango:
2 5 6   1 4 3   7 8 9
M:
13/3 = 4,33   8/3 = 2,67   24/3 = 8 GM = 5
S^2:
8,67/2 = 4,34   4,66/2 = 2,33   2/2 = 1
Punto de corte F necesario ($gl = 2, 6$; $p < 0{,}05$) = 5,14
$S^2_{entre} = (SS/gl)(n) = \{[(4{,}33 - 5)^2 + (2{,}67 - 5)^2 + (8 - 5)^2]/(3 - 1)\}(3) = (14{,}88/2)(3) = 22{,}32$
$S^2_{dentro} = (4{,}34 + 2{,}33 + 1)/3 = 2{,}56$; $F = 22{,}32/2{,}56 = 8{,}72$
Conclusión: se rechaza la hipótesis nula.

c) Explicación: comúnmente, en estos casos en los que se prueba la significación de la diferencia entre tres medias, se realizaría un análisis estándar de varianza de un criterio. Sin embargo, un supuesto del análisis de varianza establece que las poblaciones correspondientes a cada grupo están distribuidas normalmente. Según la muestra, las calificaciones dadas por el grupo que miró la película que causaba tristeza parecían muy asimétricas hacia la izquierda y, posiblemente, las calificaciones del grupo que vio la película que causaba enojo también lo fueran. (Es más, existía bastante diferencia entre las estimaciones de varianza poblacional del grupo de la película triste y del grupo de la película alegre, hecho que cuestiona otro de los supuestos del

ANOVA, que establece que las distribuciones poblacionales tienen la misma varianza).

Para resolver este problema, cambiamos cada uno de los valores observados por su rango, en todos los casos. El proceso arriba mencionado produjo el efecto de convertir la distribución de calificaciones en una distribución rectangular (aunque en realidad no ayudó mucho en cuanto al grupo de la película triste). De todas maneras, algunos estadísticos recomiendan que si los supuestos de un análisis de varianza común son cuestionables, uno debería cambiar los valores primero a rangos y luego realizar el proceso, y así se obtendrán resultados más precisos. En realidad, existen procedimientos especiales que uno puede utilizar para realizar un análisis de varianza por rangos. Pero los cálculos son matemáticamente equivalentes a los que se realizan en un análisis de varianza utilizando rangos. La única diferencia es que con el procedimiento de rango y orden existen tablas especiales que, en estos casos, son más precisas que la tabla *F*. De todos modos, los estadísticos sugieren que los resultados, al utilizar una tabla *F* común en estos casos, son una buena aproximación. Dado que nuestro resultado era claramente más extremo que el punto *F* de corte, podemos aceptar esta conclusión sin temor a equivocarnos, y rechazar la hipótesis nula.

5. a) Procedimiento: con 20 diferencias de media, la diferencia de media resultante debe ser la mayor para rechazar la hipótesis nula al nivel 0,05.

Las diferencias de media, en el orden en que se presentan los grupos en el ejercicio, son las siguientes:

4,67	4	2,67	2	3,33	2	1,33	1,33	0,67	–0,67
–4,67	–4	–2,67	–2	–3,33	–2	–1,33	–1,33	–0,67	–0,67

Las diferencias de media, ordenadas de menor (más negativa) a mayor, son las siguientes:

–4,67, –4, –3,33, –2,67, –2, –2, –2, –1,33, –1,33, –0,67, –0,67, 0,67, 0,67, 1,33, 1,33, 2, 2, 2,67, 3,33, 4, 4,67

b) Explicación: supongamos que realizar la prueba solo o frente a un amigo no implicaba ninguna diferencia. En ese caso, la razón por la cual los valores observados de las personas analizadas son mayores cuando se encuentran a solas debe de ser que la asignación aleatoria accidentalmente ubicó, dentro de la condición en la que se encuentran a solas, más personas que, de todos modos, hubieran realizado bien la tarea. Pero, ¿cuál es la probabilidad de que lo antedicho ocurra? Existen sólo 20 formas de combinar a seis personas en dos grupos de tres. Esas 20 combinaciones fueron presentadas en el enunciado del problema, y se calculó la diferencia del promedio de los valores correspondientes a las personas que realizaron la tarea a solas, menos el promedio de los valores de aquellos que realizaron la tarea en presencia de un amigo. De las 20 posibles combinaciones de asignación aleatoria, sólo una, la que presenta los valores de los dos grupos reales, habría producido semejante diferencia entre los dos grupos. Si los resultados se dieran de modo casual, existe sólo una probabilidad del 5% de obtener el mayor de 20 resultados. Ese porcentaje es demasiado bajo para considerarlo probable. Por lo tanto, se llegó a la conclusión de que la gran diferencia entre los promedios de los dos grupos no fue un hecho casual resultante de la asignación aleatoria. Dado que todos los demás aspectos entre los grupos eran iguales, la conclusión es que la situación de estar a solas o estar en presencia de un amigo es lo que ocasionó la diferencia.

6. Miller deseaba examinar la relación entre las variables que estaba analizando, probablemente incluyendo varias técnicas paramétricas de prueba de hipótesis tales como la prueba *t* o un análisis de varianza (o probando la significación de los resultados de una correlación o regresión múltiple o bivariada). Todos esos procedimientos se basan en el supuesto de que las distribuciones de las variables en la población siguen una distribución normal. Sin embargo, antes de realizar los procedimientos mencionados, Miller controló las distribuciones de varias de las variables que estaba organizando. Al realizar ese control, descubrió que las observaciones correspondientes a dos medidas clave (el índice de atención a alternativas y el tiempo transcurrido observando las diapositivas) eran positivamente asimétricos (presentaban una distribución ladeada con una larga cola hacia la derecha). Por ello resultaba poco probable que las distribuciones poblacionales de esas variables cumplieran el supuesto de seguir una curva normal. Por lo tanto, Miller decidió cambiar cada valor matemáticamente. Este proceso se denomina transformación. En este caso, calculó el logaritmo de cada registro. El efecto del proceso mencionado es reducir todos los números, pero los números mayores en mayor grado, reduciendo de ese modo la asimetría positiva y acercando la distribución a la normal. Parece particularmente adecuado realizar el tipo de trans-

formación descripta con una medida de informe propio, en la que no existe una escala absoluta. Probablemente, el mismo proceso sea apropiado también para el tiempo transcurrido observando las diapositivas, ya que cada segundo adicional de observación puede no representar una cantidad igual de interés adicional. De todos modos, es importante subrayar que una transformación de este tipo aún conserva intacto el orden de las observaciones. De todos modos, luego de realizar la transformación, los valores transformados probablemente se utilizaron en las técnicas estadísticas paramétricas comunes.

Capítulo 16

1.

gl:	5	10	15	20
t	2,571	2,228	2,132	2,086
t^2	6,61	4,96	4,55	4,35
F	6,61	4,97	4,54	4,35

2. i) ANOVA:

Punto de corte F (gl = 1, 58; p < 0,05) = 4,02

S^2_{entre} = $(SC/gl)(n)$ = ({[(12 – 11,55)^2 + (11,1 – 11,55)2]/= (2 –1)}(30) = (0,405/1)(30) = 12,15

S^2_{dentro} = (2,4 + 2,8)/2 = 2,6; F = 12,15/2,6 = 4,67

Se rechaza la hipótesis nula.

Comparación:

	gl	Punto de corte	Varianza intragrupal	t ó F
t	58	2,004	S^2_{Dentro} = 2.6	2,16
F	58	4,02	S^2_{Dentro} = 2.6	4,67
		($\sqrt{\ }$ = 2,005)		($\sqrt{\ }$ = 2,16)

ii) ANOVA:

Punto de corte F (gl = 1, 70; p < 0,05) = 3,98

S^2_{entre} = $(SC/gl)(n)$
= ({[(100 – 102)2 + (104 – 102)^2]/
= (2 –1)}(36)
= (8/1)(36) = 288

S^2_{dentro} = (40 + 48)/2 = 44; F = 288/44 = 6,55

Se rechaza la hipótesis nula.

Comparación:

	gl	Punto de corte	Varianza intragrupal	t ó F
t	70	1,995	S^2_{Dentro} = 44	2,56
F	70	3,98	S^2_{Dentro} = 44	6,55
		($\sqrt{\ }$ = 1,995)		($\sqrt{\ }$ = 2,56)

iii) Prueba t:

Punto de corte t
(gl = 30, p < 0,05, dos colas) = 2,043
$S^2_{Combinada}$ = ([15/30][8]) + ([15/30][6]) = 7;
S^2_{M1} = 7/16 = 0,44; S2M2 = 7/16 = 0,44;
$S^2_{Diferencia}$ = 0,44 + 0,44 = 0,88;
$S_{Diferencia}$ = 0,94; t = (73 – 75)/0,94 = –2,13
Se rechaza la hipótesis nula.

ANOVA:

Punto de corte F (gl = 1, 30; p < 0,05) = 4,17
S^2_{entre} = {[(73 – 74)2 + (75 – 74)2]/(2 –1)}(16)= (2/1)(16) = 32
S^2_{dentro} = (8 + 6)/2 = 7; F = 32/7 = 4,57
Se rechaza la hipótesis nula.

Comparación:

	gl	Punto de corte	Varianza intragrupal	t ó F
t	30	2,043	S^2_{Dentro} = 7	2,13
F	30	4,17	S^2_{Dentro} = 7	4,57
		($\sqrt{\ }$ = 2,042)		($\sqrt{\ }$ = 2,14)

3.

Cálculos prueba t

Punto de corte t
(gl = 18, p < 0,05, dos colas): 2,101
Diferencia media
= 170 – 150
= 20

$gl_{Total} = gl_1 + gl_2$
= 9 + 9 = 18

Cálculos ANOVA

Punto de corte F
(gl = 1, 18; p <0,05): 4,41 ($\sqrt{\ }$ = 2,1)
GM = (170 + 150)/2
= 320/2= 160
$\Sigma(M - GM)^2$ = (170 – 160)^2 + (150 – 160)^2
= 10^2 + (–10)^2
= 100 + 100= 200

S^2_{entre} ó CM_{entre} = $\dfrac{\Sigma(M - GM)^2(n)}{gl_{entre}}$
= (200/1)(10)
= 2.000

gl_{dentro}
= $gl_1 + gl_2 + ... + gl_{Último}$
= 9 + 9 = 18

$S^2_{Combinada} = (gl_1/gl_{Total})(S^2_1) + (gl_1/gl_{Total})(S^2_2)$
$= (0{,}5)(48) + (0{,}5)(32)$
$= 24 + 16 = 40$

S^2_{dentro} ó $CM_{dentro} = (S^2_1 + S^2_2 + \ldots + S^2_{Último}) / (N_{Grupos})$
$= (48 + 32) / 2 = 40$

$S^2_{Diferencia}$
$= S^2_{M1} + S^2_{M2}$
$= (S^2_{Combinada}/N_1) + (S^2_{Combinada}/N_2)$
$= (40/10) + (40/10)$
$= 4 + 4 = 8$

$S_{Diferencia}$
$= \sqrt{S^2_{Diferencia}}$
$= \sqrt{8} = 2{,}83.$

$t = (M_1 - M_2)/S_{Diferencia}$
$= 20/2{,}83$
$= 7{,}07$

Se rechaza
la hipótesis nula.

$F = S^2_{entre}/S^2_{dentro}$
ó CM_{entre}/CM_{dentro}
$= 2.000/40 = 50$
$(\sqrt{F} = 7{,}07)$

Se rechaza
la hipótesis nula.

4. $GM = (85 + 27)/8 = 14$

Grupo A

	X	$X - GM$		$X - M$		$M - GM$	
		Desv	*Desv²*	*Desv*	*Desv²*	*Desv*	*Desv²*
	13	–1	1	–4	16	3	9
	16	2	4	–1	1	3	9
	19	5	25	2	4	3	9
	18	4	16	1	1	3	9
	19	5	25	2	4	3	9
Σ	85		71		26		45
$M =$	17						

Grupo B

	X	$X - GM$		$X - M$		$M - GM$	
		Desv	*Desv²*	*Desv*	*Desv²*	*Desv*	*Desv²*
	11	–3	9	2	4	–5	25
	7	–7	49	–2	4	–5	25
	9	–5	25	0	0	–5	25
Σ	27		83		8		75
$M =$	9						

Nota: Desv = Desvío; Desv² = Desvío cuadrático

$\Sigma(X - GM)^2$ ó $SC_{Total} = 71 + 83 = 154$
$\Sigma(X - M)^2$ ó $SC_{dentro} = 26 + 8 = 34$
$\Sigma(M - GM)^2$ ó $SC_{entre} = 45 + 75 = 120$
Control ($SS_{Total} = SC_{dentro} + SC_{entre}$): $154 = 34 + 120$

Grados de libertad:
$gl_{Total} = N - 1 = 8 - 1 = 7$
$gl_{dentro} = gl_1 + gl_2 + \ldots + gl_{Último} = 4 + 2 = 6$
$gl_{entre} = N_{Grupos} - 1 = 2 - 1 = 1$
Control ($gl_{Total} = gl_{dentro} + gl_{entre}$): $7 = 6 + 1$

Estimaciones de varianza poblacional:
S^2_{dentro} ó $CM_{dentro} = SC_{dentro}/gl_{dentro} = 34/6 = 5{,}67$
S^2_{entre} ó $CM_{entre} = SC_{entre}/gl_{entre} = 120/1 = 120$
Razón F: $F = S^2_{entre}/S^2_{dentro}$ ó CM_{entre}/CM_{dentro}
$= 120/5{,}67 = 21{,}16$

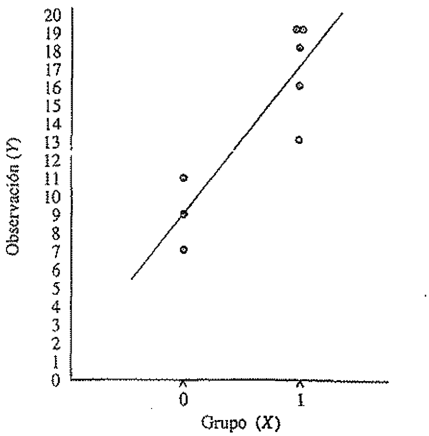
Proporción de varianza explicada:
$R^2 = SC_{entre}/SS_{Total} = 120/154 = 0{,}78$

Correlación (Grupo A = 1, Grupo B = 0):

Grupo (X) Observación (Y)

X		Y					
Original	*Z*	*Original*	*Z*	Z_XZ_Y	$\hat{Y}$	*Error*	*Error²*
1	0,77	13	–0,22	–0,17	17	–4	16
1	0,77	16	0,46	0,35	17	–1	1
1	0,77	19	1,14	0,88	17	2	4
1	0,77	18	0,92	0,71	17	1	1
1	0,77	19	1,14	0,88	17	2	4
0	–1,29	11	–0,68	0,88	9	2	4
0	–1,29	7	–1,60	2,06	9	–2	4
0	–1,29	9	–1,14	1,47	9	0	0
Σ:5		112		7,06			34
$M = 0{,}625$		14	$r = 0{,}88$ $r^2 = 0{,}77$				
$SC = 1{,}874$		154					
$SD = 0{,}484$		4,387					

Reducción proporcional del error $= r^2$
= reducción del error/error total
$= (SC_{Total} - SC_{Error})/SC_{Total}$
$= (154 - 34)/154$
$= 120/154 = 0{,}78$



Similitudes:

r	F
Media de $Y = 14$	Gran media $= 14$
$SC_{\text{Total}} = 154$	$SC_{\text{Total}} = 154$
Y predicha	
para Grupo A = 17	Media de grupo A = 17
Y predicha	
para Grupo B = 9	Media de grupo B = 9
$SC_{\text{Error}} = 34$	$SC_{\text{dentro}} = 34$
Reducción de error = 120	$SC_{\text{entre}} = 120$
$r^2 = 0{,}77$	$R^2 = 0{,}78$

5. Prueba t:

Punto de corte t ($gl = 4$, $p < 0{,}05$, una cola) $= 2{,}132$

Grupo A: $M = 0{,}8$, $S^2 = 0{,}01$; Grupo B: $M = 0{,}4$, $S2 = 0{,}04$

$S^2_{\text{Combinada}} = ([2/4][0{,}01]) + (2/4][0{,}04]) = 0{,}025$; $S^2_{\text{MA}} = 0{,}025/3 = 0{,}0083$; $S^2_{\text{entre}} = 0{,}0083$;

$S^2_{\text{Diferencia}} = 0{,}0083 + 0{,}0083 = 0{,}017$; $S_{\text{Diferencia}} = 0{,}13$; $t = (0{,}8 - 0{,}4)/0{,}13 = 3{,}08$

Se rechaza la hipótesis nula.

Correlación (Grupo A = 1, Grupo B = 0):

Grupo (X)		Observación (Y)		
Original	Z_X	*Original*	Z_Y	$Z_X Z_Y$
1	1	0,7	0,42	0,42
1	1	0,9	1,26	1,26
1	1	0,8	0,84	0,84
0	−1	0,6	0,00	0.00
0	−1	0,4	−0,84	0,84
0	−1	0,2	−1,68	1,68
Σ 3		3,6		5,04
$M = 0{,}5$		0,6	$r = 0{,}84$	
$SD = 0{,}5$		0,238		

$t = r\sqrt{N-2}/\sqrt{1-r^2} = 0{,}84\sqrt{6-2}/\sqrt{1-0{,}84^2} = 0{,}84\sqrt{4}/\sqrt{1-0{,}71}$
$= (0{,}84)(2)/\sqrt{0{,}29} = 1{,}68/0{,}54 = 3{,}11$

6. El análisis de varianza se puede considerar como el estudio de la relación entre la variable independiente (la variable en la que difieren los grupos, tal como la "condición experimental") y una variable dependiente. En este sentido, es igual a la correlación y regresión, las cuales también estudian la relación entre una variable dependiente y una independiente. De hecho, supongamos que en un análisis de varianza de dos grupos codificáramos a las personas en uno de los grupos con el número 1 y a las del otro grupo con el 2 (o con dos números cualesquiera, los utilizados son sólo ejemplos). Si después calculáramos la correlación entre ese número de pertenencia al grupo y la variable dependiente, la significación de la correlación será igual a la significación del análisis de varianza.

Existe una variedad de conexiones entre los dos tipos de cálculo. Primero, tanto en la regresión como en el análisis de varianza, se calculan los desvíos cuadráticos totales con respecto a la media general. En ambos, el proceso se denomina suma de cuadrados total. Otro vínculo más profundo surge debido a que la mejor predicción para los integrantes de un grupo es la media de ese grupo. Así, la ecuación de regresión predice la media para los miembros de cada grupo. El resultado es que los errores de las predicciones son desvíos de las observaciones con respecto a la media. Si los elevamos al cuadrado y los sumamos, se denominan suma de cuadrados del error en la regresión, y "suma de cuadrados intragrupal" en un análisis de varianza. En la regresión, como paso preliminar para calcular la reducción proporcional de error, calculamos la reducción de error ($SC_{\text{Total}} - SC_{\text{Error}}$). Es la cantidad de error cuadrático que la fórmula de regresión evita con respecto a predecir a partir de la media de la variable dependiente. La cantidad calculada del modo descripto resulta ser la misma que la "suma de cuadrados intergrupales" en un análisis de varianza. Lo anterior se debe a que cuando hay sólo dos medias grupales involucradas, la regresión puede mejorar la predicción sólo en la medida en que difieren las medias de los dos grupos. Finalmente, dado que SC_{Total} es igual en la regresión y en el análisis de varianza, y la reducción de error es igual a SC_{entre}, por consiguiente r^2 en la regresión será igual a R^2 calculado como un tamaño de efecto en el análisis de varianza.

Sin embargo, existe una complicación en cuanto a este vínculo entre la regresión y el análisis de varianza. La regresión, incluso la regresión múltiple, utiliza sólo variables numéricas ordinarias. Cuando existen sólo dos grupos, codificándolos con dos números creamos una variable numérica que funciona adecuadamente. Pero cuando hay tres o más grupos, la variable nominal según la cual difieren esos grupos no se puede utilizar directamente en el análisis de regresión. La solución es crear más de una de esas variables numéricas con dos valores para que cumplan el papel de la variable nominal de predicción. Luego, una correlación múltiple realizada utilizando las variables especialmente codificadas arroja el mismo resultado que un análisis de varianza (en términos de significación estadística).

Capítulo 17

1. Una regresión múltiple jerárquica es una variante de la regresión múltiple común, en la cual se agregan, de a una por vez, cada una de las variables de predicción a la regla de predicción (a veces se agrega una serie de variables como grupo), y se calcula el aporte adicional de esa variable (además de la variable agregada en el paso anterior). El investigador determina de antemano el orden en el que se ingresan las variables de predicción. En el estudio de Lindzey et al., la variable dependiente era la capacidad de los niños para adaptarse a la vida social. Las dos primeras variables de predicción que se tuvieron en cuenta fueron los índices de iniciativa de padre e hijo, que justificaron sólo el 3% de la varianza en la variable dependiente. Es decir, la R^2 fue de 0,03. (La R^2 no fue significativa, los investigadores observaron que el nivel de significación estaba lejos de ser menor a 0,05, de hecho era un muy alto 0,57). Hasta aquí, el procedimiento es similar a una regresión múltiple común con dos variables de predicción y una variable dependiente.

   Sin embargo, luego los investigadores agregaron una variable de predicción adicional, el *cumplimiento recíproco entre padre e hijo*. Lindzey et al. informan que la varianza general explicada aumentó un 18%, lo que significa que la R^2 tuvo que haber crecido de 0,03 a 0,21 (es decir, 3% + 18% = 21%). Más aún, los investigadores observan que el aumento del 18% era significativo (con un nivel $p < 0{,}01$).

   El proceso descriptivo nos indica que los índices de iniciativa de padre e hijo no son muy importantes en la predicción de la capacidad del niño y, lo que es más importante, aun teniendo en cuenta esas dos variables el cumplimiento recíproco realiza un gran aporte a la predicción de la capacidad del niño para adaptarse a la vida social.
2. Boyd y Gullone están describiendo el grado de confiabilidad de las medidas que utilizan en su estudio. La confiabilidad es el grado de coherencia con que la prueba mide determinado aspecto, es decir, en qué medida se obtendrían los mismos resultados si las mismas personas volvieran a realizar la prueba en idénticas circunstancias. Una manera de evaluar la confiabilidad es observando la correlación entre una y otra mitad de la prueba –la idea es que la misma persona está realizando dos pruebas (las dos mitades de la prueba) al mismo tiempo bajo las mismas circunstancias. El alfa de Cronbach es una medida común de confiabilidad, es a lo que se refieren los investigadores cuando mencionan el "coeficiente alfa" al discutir la confiabilidad. Para ser más precisos, el alfa de Cronbach indica el promedio general de correlaciones entre cada posible división de la prueba en mitades, y luego adapta el cálculo de modo tal de tener en cuenta el hecho de que, con sólo la mitad de los ítems, las correlaciones son un poco menores que si la prueba completa se correlacionara con otra prueba completa similar. Por lo general, un alfa de Cronbach de 0,60 ó 0,70 se considera un nivel mínimo adecuado de confiabilidad. El hecho de que algunas de las medidas de Boyd y Gullone fueran menores a los valores indicados significa que algunas de las variables estudiadas pueden no estar proporcionando información muy precisa. En realidad, las correlaciones entre medidas con bajas confiabilidades pueden subestimar la verdadera correlación entre las variables medidas.
3. Un análisis factorial ayuda al investigador que ha medido a los participantes con respecto a una gran cantidad de variables, a descubrir el patrón implícito (si existe) entre ellas, es decir, a descubrir qué variables se agrupan en el sentido de correlacionarse entre sí, pero no con variables que no pertenecen al grupo. En el estudio de Fawzi et al., los investigadores contaban con puntuaciones en 16 síntomas de PTSD tomadas de entrevistas a 74 refugiados vietnamitas. Los resultados del análisis factorial sugieren que el patrón implícito más adecuado presenta cuatro agrupaciones o factores

(que en la tabla de Fawzi et al. se denominan "dimensiones"). La tabla indica las correlaciones, llamadas "cargas factoriales", de cada variable individual con el grupo.

Cuando los investigadores diseñaron la tabla, sólo incluyeron la carga factorial de cada síntoma en el factor en el cual presentaba la mayor carga. (Cada variable de un análisis factorial presenta una carga en cada factor, pero comúnmente presenta una carga alta sólo en un factor y, por lo tanto, se considera parte de ese factor). Los investigadores observan que los primeros tres factores corresponden a los tres aspectos clave del PTSD, en la forma en que este se entiende comúnmente. Sin embargo, el cuarto factor (que incluye sólo un ítem) sugería que existe un aspecto de evación adicional, y de cierto modo independiente, que no había sido considerado como tal en trabajos previos.

4. a) En el contexto del diseño propuesto, el resultado clave del sendero hipotético, desde "calidad de deseable" a "intensidad", es que ese *sendero fue significativo en los tres estilos de* vinculación, aunque fue claramente más fuerte en el caso de aquellos con estilo de vinculación ansiosa-ambivalente. El sendero hipotético de la probabilidad fue bajo para los tres estilos de vinculación, aunque fue significativo para los evasivos. Finalmente, el sendero hipotético desde el "deseo de estar enamorado" fue positivo en el caso de los evasivos (es decir, que a mayor deseo de estar enamorado, mayor era la intensidad), pero el mismo sendero resultó negativo en el caso de los ansiosos-ambivalentes (es decir, que a mayor deseo de estar enamorados, menor era la intensidad); y en el caso de los seguros, el sendero presentó muy poca relación.

   b) El diseño de ecuación estructural es una técnica estadística en la que el investigador especifica un patrón de relaciones causales entre las variables, diagramado con flechas que conectan cada causa con su efecto. Como parte del proceso, los investigadores también pueden especificar que, en realidad, algunas variables medidas en el estudio son efectivamente indicadores de una variable "latente" implícita no medida. En este ejemplo, para cada grupo, el investigador ha especificado senderos desde los tres factores de motivación hacia la intensidad. Aun más, cada una de las tres variables más importantes (representadas en óvalos) son, en efecto, variables latentes que se revelan a través de diversas variables medidas (representadas en este caso por flechas que parten desde cada óvalo –si no hubiera sido por la falta de espacio esas flechas se dirigirían cada una a un rectángulo que indicaría una variable medida específica).

   Un aspecto estadístico clave del diseño de ecuación estructural involucra la utilización de correlaciones entre variables para calcular un "coeficiente de senderos" para cada flecha. El coeficiente de senderos indica el grado en el cual los cambios de la variable ubicada en la base de la flecha están relacionados con los cambios en la variable ubicada en la punta de la flecha (bajo condiciones en las cuales todas las otras causas de esa variable efecto se mantienen constantes). Es decir, el coeficiente de senderos es un coeficiente de regresión estandarizado (un "beta") de la variable causal, en una ecuación en la que la variable efecto es la variable dependiente y todas las variables causales son variables de predicción. Por ejemplo, en el caso de los seguros, el sendero de 0,42 desde "calidad de deseable" hacia "intensidad" significa que, manteniendo constante la probabilidad y el deseo del estado, por cada desvío estándar de cambio en la calidad de deseable se produciría un desvío estándar de 0,42 en el cambio de intensidad.

5. El "análisis de varianza multivariado", descripto en esta publicación, es igual a un análisis de varianza 2 x 2 común, excepto por el hecho de que en un análisis multivariado se incluyen diversas variables dependientes al mismo tiempo. En este ejemplo, había cinco medidas de preferencia en cuanto a la forma de resolución de conflictos. El resultado significativo del "efecto principal de la cultura" significa que las dos culturas presentaron una diferencia significativa al tomarse en cuenta de una sola vez toda la serie de variables dependientes. En forma similar, el resultado significativo de "cultura por tipo de interacción conflictiva" significa que el efecto de la cultura en la serie de variables dependientes varía según el tipo de conflicto. Para comprender cuál de las diversas variables de resolución de conflictos justificaba los efectos generales, los investigadores realizaron análisis de varianzas comunes con una variable dependiente por vez (a esto se refieren cuando mencionan un "análisis univariado"). Como consecuencia, los patrones de resultados fueron muy diferentes según la variable dependiente específica que se tomaba en cuenta.

6. a) Diseño causal (análisis de senderos o diseño de variable latente).
   b) Correlación y regresión bivariada.

c) Cálculo estadístico de confiabilidad, tal como el alfa de Cronbach y la confiabilidad por prueba y reprueba.
d) Análisis de varianza multivariado 3 x 2, probablemente seguido de análisis de varianza univariado 3 x 2, e incluso también por comparaciones múltiples univariadas y/o multivariadas entre pares o grupos específicos de medias.
e) Regresión por pasos.
f) Análisis factorial.
g) Prueba *t* para medias independientes.
h) Análisis de varianza de un criterio, posiblemente seguido de comparaciones múltiples entre pares o grupos específicos de medias.
i) Regresión múltiple jerárquica.

Glosario

Los números en paréntesis se refieren a los capítulos en los que el término fue presentado o tratado sustancialmente.

Alfa (α): probabilidad de cometer un error Tipo I; es igual al **nivel de significación** (8). También es la forma corta de referirse al **alfa de Cronbach**. (17)

Alfa de Cronbach's (α): índice de confiabilidad de medidas ampliamente utilizado, que equivale al promedio de las correlaciones por mitades de todas las posibles divisiones en mitades de los ítems de una prueba. (17)

Análisis causal: procedimiento, tal como el análisis de sendero o el modelo de ecuación estructural, que analiza correlaciones entre un grupo de variables en función de un patrón predicho de relaciones causales entre ellas. (17)

Análisis de covarianza (ANCOVA): análisis de varianza que se realiza después de adaptar las variables para controlar el efecto de una o más variables adicionales no deseadas. (17)

Análisis de senderos: método de análisis de las correlaciones entre un grupo de variables según un patrón predicho de relaciones causales; usualmente el patrón predicho se diagrama en forma de un patrón de flechas que van desde las causas hacia los efectos. (17)

Análisis de varianza de cuadrados mínimos: método recomendado para el análisis factorial de varianza cuando en las distintas casillas hay cantidades desiguales de participantes. (13)

Análisis de varianza de dos criterios: análisis de varianza para un diseño factorial eden dos crieterios. (13)

Análisis de varianza de medidas repetidas: análisis de varianza en el que se mide a cada individuo más de una vez, de forma tal que los niveles de la(s) variable(s) independiente(s) están conformados por las diferentes ocasiones o los distintos tipos de prueba aplicadas a las mismas personas. (13)

Análisis de varianza de un criterio: análisis de varianza en el que existe sólo una variable independiente. (11, 12)

Análisis factorial de varianza: análisis de varianza para un diseño factorial de investigación; análisis de varianza según las diferencias entre las medias de los niveles de cada variable independiente y según la interacción de las variables independientes. (13)

Análisis factorial: procedimiento estadístico aplicado en situaciones en las que se miden muchas variables, que identifica agrupaciones de variables que se correlacionan al máximo entre sí y en forma mínima con otras variables. (17)

(ANOVA) **Análisis de varianza:** procedimiento de prueba de hipótesis para estudios que incluyen dos o más grupos. (11–13)

Asimetría: grado en el cual una distribución de frecuencias presenta más casos de un lado de su punto medio, como contrapartida de la distribución perfectamente simétrica. (1)

Beta (β): coeficiente de regresión estandarizado. (4) Es también la posibilidad de cometer un error Tipo II en la prueba de hipótesis. (8)

Casilla: en un diseño factorial, es la combinación particular de niveles de las variables independientes. (13) En la prueba chi-cuadrado, es la combinación particular de categorías de dos variables en una tabla de contingencia. (14)

Caso aislado: registro con un valor extremo (muy alto o muy bajo) en relación con el resto de los registros de la *distribución*. (2)

Codificación nominal: conversión de una variable de predicción nominal (categórica), de un análisis de varianza, en diversas variables numéricas de dos niveles que pueden utilizarse en un análisis de regresión múltiple. (16)

Coeficiente de correlación (r): promedio de los productos cruzados de las puntuaciones Z de dos variables; medida del grado de correlación lineal que va del –1 (correlación lineal negativa perfecta), pasando por el 0 (ausencia de correlación) hasta el +1 (correlación lineal positiva perfecta); raíz cuadrada de la reducción proporcional de error. (3, 4)

Coeficiente de correlación múltiple (R): una medida de la asociación general entre una variable dependiente y la combinación de dos o más variables de predicción; la raíz cuadrada positiva de la reducción proporcional de error (R^2) en un análisis de regresión múltiple. (4)

Coeficiente de correlación parcial: correlación entre dos variables por encima de la influencia de otra u otras variables. (17)

Coeficiente de regresión estandarizado (β): coeficiente de regresión en una norma de predicción que utiliza puntuaciones Z; también se lo denomina "valor ponderado beta". (4)

Coeficiente de regresión para puntuaciones originales (b): el coeficiente de regresión en un modelo de predicción (ecuación de regresión) que utiliza puntuaciones originales. (4)

Coeficiente de regresión (b, β): el número que se multiplica por el registro de una persona en la variable independiente, como parte de una fórmula (norma de predicción) para predecir registros en la variable dependiente. (4)

Coeficiente de senderos: grado de asociación relacionado con una flecha en un análisis de senderos (que incluye un modelo de ecuación estructural); es igual a un coeficiente de regresión estandarizado de una norma de predicción de regresión múltiple, en el cual la variable, al final de la flecha, es la variable dependiente y la variable al comienzo de la flecha es la variable de predicción (siendo también variables de predicción todas las demás variables que tienen flechas que van hacia la variable dependiente). (17)

Coeficiente phi (ϕ): medida de la relación entre dos variables nominales dicotómicas, equivalente a una correlación de las dos variables si se les adjudicaran valores numéricos (por ejemplo 1 y 0 para las dos categorías); medida del tamaño de efecto para la prueba chi-cuadrado de independencia con tabla de contingencia 2 x 2. (14)

Comparaciones múltiples: procedimientos de prueba de hipótesis para analizar las diferencias entre determinadas medias en el contexto de un análisis de varianza general. (12)

Comparaciones planificadas: comparaciones múltiples en las que se designan de antemano las medias particulares a comparar; es igual a los contrastes planificados. (12)

Comparaciones *post hoc*: comparaciones múltiples entre determinadas medias no designadas de antemano; procedimiento realizado como parte de un análisis de exploración después de haber llevado a cabo el estudio. (12)

Confiabilidad por división en mitades: es un índice de la confiabilidad de las medidas que se basa en la correlación de los registros correspondientes a los ítems de las dos mitades de la prueba. (17)

Confiabilidad por prueba y repetición: un índice de la confiabilidad de las medidas, obtenido a través de administrar la medida dos veces a un grupo de personas; es la correlación entre los registros obtenidos en las dos pruebas. (17)

Confiabilidad: grado de coherencia de una medida; indica hasta qué punto, si aplicáramos la misma medida nuevamente a la misma persona bajo las mismas circunstancias, obtendríamos el mismo resultado. (3, 17)

Constante de regresión (a): un número fijo determinado que se agrega a la predicción en un modelo de predicción (ecuación de regresión) que utiliza puntuaciones originales. (4)

Contraste lineal: comparación planificada en la que cada uno de los diferentes niveles de la variable independiente tiene valores numéricos significativos; se asemeja a una correlación en la que por cada participante, una variable es la influencia predicha del grupo en el que se encuentra el participante y la otra variable es el registro con respecto a aquello con que se lo está midiendo. (12)

Contrastes planificados: es igual a las **comparaciones planificadas.** (12)

Controlar: es la anulación de la influencia de una variable en la asociación entre las otras variables en la regresión múltiple, la correlación parcial o el análisis de covarianza; es lo mismo que **excluir o mantener constante.** (17)

Corrección por atenuación: procedimiento estadístico que calcula la correlación que se esperaría entre dos variables si ambas variables fueran medidas con una confiabilidad perfecta. (3)

Correlación curvilínea: una relación entre dos variables que se refleja en un diagrama de dispersión en forma de puntos que siguen un patrón sistemático distinto de una línea recta; toda asociación entre dos variables distinta de una correlación lineal.(3)

Correlación lineal: una relación entre dos variables que se refleja en un diagrama de dispersión en forma de puntos que siguen una línea recta; correlación en la que *r* es distinta de 0. (3)

Correlación negativa: una relación entre dos variables en la que los registros altos de una coinciden con los bajos de la otra, los medios con los medios y los bajos con los altos; en un diagrama de dispersión, los puntos siguen aproximadamente una línea recta que se inclina hacia abajo y hacia la derecha; una correlación en la que *r* es menor que 0. (3)

Correlación nula: ausencia de relación sistemática entre dos variables. (3)

Correlación perfecta: una relación entre dos variables que se refleja en un diagrama de dispersión en forma de puntos que siguen una línea recta en forma exacta; una correlación en la que $r = 1$ ó -1; situación en la que la puntuación *Z* de cada persona en una variable es exactamente igual a la puntuación *Z* de esa misma persona en la otra variable. (3)

Correlación positiva: una correlación entre dos variables en la que los registros altos de una coinciden con los registros altos de la otra, los medios con los medios y los bajos con los bajos; en un diagrama de dispersión, los puntos siguen aproximadamente una línea recta que se inclina hacia arriba y hacia la derecha; una correlación en la que *r* es mayor que 0.(3)

Covariable: variable controlada en un análisis de covarianza. (17)

Cuadrados medios intergrupales (CM_{Entre}): es igual a la estimación intergrupal de varianza poblacional (S^2_{Entre}). (11)

Cuadrados medios intragrupales (CM_{Dentro}): es igual a la estimación intragrupal de varianza poblacional (S^2_{Dentro}). (11)

Curtosis: grado en el que una distribución de frecuencias se desvía de una curva normal, con colas que son muy espesas o muy delgadas. (1)

Curva normal: distribución de frecuencias específica, matemáticamente definida, con forma de campana, simétrica y unimodal; distribuciones observadas en la naturaleza y a las cuales se aproximan las investigaciones psicológicas. (1, 5)

Dato estadístico chi-cuadrado (χ^2): dato estadístico que refleja la ausencia general de ajuste entre las frecuencias esperadas y observadas; es la suma, teniendo en cuenta todas las categorías o casillas, de las diferencias cuadráticas entre frecuencias observadas y esperadas, dividida por la frecuencia esperada. (14)

Dato estadístico muestral: dato estadístico descriptivo, como la media o el desvío estándar, calculado a partir de los registros de un determinado grupo de personas analizadas. Los datos estadísticos muestrales generalmente se simbolizan con letras comunes (en contraposición con las griegas). (5)

Desvío cuadrático: El cuadrado de la diferencia entre un registro y la media. (2)

Desvío estándar (*SD*, *S*, σ): raíz cuadrada del promedio de los desvíos cuadráticos de la media; es el más común de los datos estadísticos descriptivos de la variación; es aproximadamente (no exactamente) la cantidad promedio de variación de los registros de una distribución con respecto a la media. (2, 5, 9)

Desvío estándar de una distribución de medias (σ_M, S_M): raíz cuadrada de la varianza de la distribución de medias; es igual al error estándar. (7, 9)

Desvío estándar poblacional (σ): desvío estándar de la población (usualmente desconocido). (5)

Desvío: un registro menos la media de todos los registros de esa distribución. (2)

Diagrama de dispersión: gráfico que refleja la relación entre dos variables; los valores de la variable independiente o de predicción se encuentran en el eje horizontal; los valores de la variable dependiente se encuentran en el eje vertical, y cada registro es representado gráficamente por un punto en este espacio bidimensional. (3)

Dimensión: en un diseño factorial, una de las variables independientes que se cruza con otra variable independiente. (13)

Diseño de medidas repetidas: estrategia de investigación en la que se prueba a cada persona más de una vez; es igual al **diseño intra-sujeto**. (9, 13)

Diseño factorial de investigación: modo de organizar un estudio en el que la influencia de dos variables o más se estudia de una sola vez preparando la situación de manera tal de probar a un grupo de personas según cada combinación de los niveles de las variables; por ejemplo, en un diseño factorial de investigación 2 x 2 habría cuatro grupos: aquellos con registro alto en la variable 1 y registro alto en la variable 2; aquellos con registro alto en la variable 1 pero bajo en la variable 2; aquellos con registro alto en la variable 2 pero bajo en la variable 1; aquellos con registro bajo en la variable 1 y registro bajo en la variable 2. (13)

Diseño factorial en dos sentidos: diseño factorial con dos variables independientes. (13)

Diseño intra-sujeto: es igual al **diseño de medidas repetidas**. (9)

Distribución bimodal: distribución de frecuencias con dos frecuencias aproximadamente iguales, cada una de las cuales es claramente mayor a cualquiera de las demás. (1)

Distribución chi-cuadrado: curva matemáticamente definida que se utiliza como distribución comparativa en las pruebas chi-cuadrado. Es la *distribución* de los datos estadísticos chi-cuadrado. (14)

Distribución comparativa: distribución que representa la situación de la población si la hipótesis nula es verdadera; la distribución con la cual se compara una estadística muestral en la prueba de hipótesis. (6)

Distribución de diferencias entre medias: distribución de todas las diferencias posibles entre medias de dos muestras, de forma tal que en cada par de medias, una pertenece a una población y la otra a una segunda población; es la distribución comparativa en una prueba *t* para medias independientes. (10)

Distribución de frecuencias: el patrón de frecuencias en los distintos valores; es lo que se describe a través de una tabla de frecuencias, un histograma o un polígono de frecuencias. (1)

Distribución de medias: distribución de todas las posibles medias muestrales de determinado tamaño seleccionadas de una población en particular (también denominada distribución de muestreo de la media); es la distribución comparativa en las pruebas de hipótesis que involucran una muestra de más de un registro. (7)

Distribución *F*: curva matemáticamente definida que describe la distribución comparativa utilizada en un análisis de varianza; distribución de las razones *F* cuando la hipótesis nula es verdadera. (11)

Distribución multimodal: distribución de frecuencias con dos o más frecuencias altas separadas por una frecuencia menor; una distribución bimodal es el caso especial en el que existen dos frecuencias altas. (1)

Distribución normal: distribución de frecuencias que sigue una curva normal. (5)

Distribución rectangular: distribución de frecuencias en la cual todos los valores tienen aproximadamente la misma frecuencia. (1)

Distribución simétrica: distribución en la cual los patrones de frecuencias a la derecha y a la izquierda son imágenes especulares entre sí. (1)

Distribución *t*: curva matemáticamente definida que describe la distribución comparativa en una prueba *t*. (9)

Distribución unimodal: distribución de frecuencias con un valor que claramente presenta una frecuencia mayor a la de cualquier otro. (1)

Efecto interactivo: situación que se presenta en el análisis factorial de varianza, en la cual la combinación de variables tiene un efecto especial que no podría predecirse a partir del conocimiento de los efectos de las dos variables en forma individual. (13)

Efecto piso: situación en la cual muchos registros se amontonan en el extremo más bajo de la escala, ya que no es posible que exista un registro menor. (1)

Efecto principal: diferencia entre los grupos de una dimensión de un diseño factorial; el resultado de una variable, promediando las divisiones en la otra u otras variables (a veces utilizada sólo para diferencias significativas). (13)

Efecto techo: Aquella situación en la cual muchos registros se acumulan en el extremo más alto de la escala creando asimetría, ya que no es posible que exista un registro mayor. (1)

Error estándar: es igual al **desvío estándar de una distribución de medias.** (7)

Error Tipo I: rechazar la hipótesis nula cuando en realidad es verdadera; obtener un resultado *estadísticamente significativo* cuando, de hecho, la hipótesis de investigación no es verdadera. (8)

Error Tipo II: no rechazar la hipótesis nula cuando en realidad es falsa; no obtener un resultado *estadísticamente significativo* cuando, de hecho, la hipótesis nula es verdadera.

Error: en la predicción, es el registro real menos el registro predicho. (4)

Estadística descriptiva: Procedimientos para resumir una serie de registros o hacerlos más comprensibles de algún otro modo. (1)

Estadística inductiva: procedimientos utilizados para sacar conclusiones sobre la base de los registros recolectados en una investigación científica (registros muestrales) pero que los exceden (creando conclusiones sobre poblaciones). (1)

Estadísticamente significativo: conclusión que establece que los resultados de un estudio serían improbables si, en efecto, no hubiera diferencia en las poblaciones representadas por las muestras analizadas; resultado de una prueba de hipótesis en la que se rechaza la hipótesis nula. (3, 6)

Estimación combinada de la varianza poblacional: ($S^2_{\text{Combinada}}$): en una prueba *t* para medias independientes, es un promedio ponderado de las estimaciones de varianza poblacional calculadas a partir de las dos muestras (cada estimación ponderada según la proporción de grados de libertad de su muestra, dividida por los grados de libertad totales de ambas muestras). (10)

Estimación intergrupal de la varianza poblacional (S^2_{Entre}, CM_{Entre}): en un análisis de varianza, es la estimación de la varianza de la distribución poblacional de individuos que se basa en la variación entre las medias de los grupos analizados; es igual a los **cuadrados medios intergrupales.** (11)

Estimación intervalar: rango de registros (es decir, registros que se encuentran entre algún valor superior e inferior específico) que se estima incluye un parámetro poblacional y se contrapone a la **estimación puntual**; un **intervalo de confianza** es un ejemplo de estimación intervalar. (7)

Estimación intragrupal de varianza poblacional (S^2_{Dentro}, CM_{Dentro}): en un análisis de varianza, es la estimación de la varianza de la distribución poblacional de individuos basada en la variación entre los registros dentro de cada uno de los grupos analizados; también se la llama cuadrados medios intragrupales; es igual a la varianza del error (S^2_{Error}) y error de los cuadrados medios (CM_{Error}). (11)

Estimación no sesgada de la varianza poblacional (S^2): estimación de la varianza poblacional basada en registros muestrales, que ha sido corregida (dividiendo la suma de desvíos cuadráticos por el tamaño de la muestra menos 1, en lugar de utilizar el procedimiento usual de dividir directamente por el tamaño de la muestra) de manera que resulte igualmente probable sobrestimar o subestimar la verdadera varianza poblacional. (2, 9)

Estimación sesgada: estimación de un parámetro poblacional que probablemente sobrestime o subestime sistemáticamente el verdadero valor poblacional. Por ejemplo, SD^2 sería una estimación sesgada de la varianza poblacional (la subestimaría sistemáticamente). (9)

Excluir: es el hecho de anular la influencia de una variable en la asociación entre las otras variables en la regresión múltiple, la correlación parcial o el análisis de covarianza; es igual a mantener **constante o controlar.** (17)

Factor: en un análisis factorial, es la sub-serie de variables correlacionadas entre sí pero no con variables fuera de la sub-serie. (17)

Forma de una distribución de medias: contorno del histograma de una distribución de medias que puede seguir una curva normal o ser asimétrico; en general, una distribución de medias tendrá tendencia a ser unimodal y simétrica, y usualmente es normal. (7)

Fórmula de cálculo: ecuación matemáticamente equivalente a la fórmula de definición que es más fácil de utilizar para cálculos manuales pero que no muestra directamente el significado del procedimiento que simboliza. (2)

Fórmula de definición: ecuación que muestra directamente el significado del procedimiento que simboliza. (2)

Fórmula de predicción con puntuaciones originales: norma de predicción (ecuación de regresión) que utiliza puntuaciones originales. (4)

Frecuencia esperada: en una prueba chi-cuadrado, es la cantidad esperada de individuos en una categoría o casilla si la hipótesis nula fuera verdadera. (14)

Frecuencia marginal: en una prueba chi-cuadrado, la frecuencia (cantidad de casos) en una fila o columna de una tabla de contingencia. (14)

Frecuencia observada: en una prueba chi-cuadrado, es la cantidad de individuos en una categoría o casilla efectivamente encontrados a través de un estudio. (14)

Frecuencia relativa esperada: la cantidad de resultados exitosos dividida por la cantidad de resultados totales que se esperaría obtener si se repitiera un experimento una gran cantidad de veces. (5)

Grado de correlación: la medida en la cual existe un patrón claro de alguna relación en particular (generalmente lineal) entre dos variables. (3)

Grados de libertad (gl): cantidad de registros libres para variar cuando se estima un parámetro poblacional; comúnmente parte de una ecuación que se utiliza para realizar esa estimación. Por ejemplo, en la fórmula para la estimación de la varianza poblacional a partir de una sola muestra, los grados de libertad son la cantidad de registros menos 1. (9–14)

Grados de libertad del denominado r (gl_{Dentro}): grados de libertad utilizados en la estimación intragrupal de varianza en un análisis de varianza; forman el denominador de la razón F; la cantidad de registros libres para variar (cantidad de registros de cada grupo menos 1, sumando los de todos los grupos) en el cálculo de la estimación intragrupal de varianza; grados de libertad intragrupales. (11)

Grados de libertad del numerador (gl_{Entre}): grados de libertad utilizados en la estimación intergrupal de varianza en un análisis de varianza (el numerador de lal razón F); cantidad de registros libres para variar (cantidad de grupos menos 1) en el cálculo de la estimación intergrupal de varianza; grados de libertad intergrupales. (11)

Grados de libertad intergrupales (gl_{Entre}): es igual a los **grados de libertad del numerador**. (11)

Grados de libertad intragrupales (gl_{Dentro}): es igual a los **grados de libertad del denominador**. (11)

Hipótesis de investigación: en la prueba de hipótesis, afirmación acerca de la relación predicha entre poblaciones (comúnmente es una predicción de diferencias entre medias poblacionales). (6)

Hipótesis direccional: hipótesis de investigación que predice una determinada dirección de la diferencia entre poblaciones; por ejemplo, que una población tendrá una media mayor que la otra población. (6)

Hipótesis no direccional: hipótesis de investigación que no predice ninguna dirección determinada de la diferencia entre poblaciones. (6)

Hipótesis nula: afirmación sobre una relación entre poblaciones que representa la noción crucial opuesta a la hipótesis de investigación; afirmación que establece que en la población no existe diferencia (o existe una diferencia opuesta a la predicha) entre poblaciones; una afirmación artificial establecida para analizar si puede ser rechazada como parte del proceso de prueba de hipótesis. (6)

Histograma: tipo de gráfico de barras de una distribución en el cual los valores se marcan en el eje horizontal y la altura de cada barra es la frecuencia de ese valor; las barras se ubican una al lado de la otra sin espacios intermedios, dando la apariencia del contorno de una ciudad en el horizonte. (1)

Independencia: situación en la que no existe relación sistemática entre dos variables; término utilizado generalmente en relación con dos variables nominales en el contexto de una prueba chi-cuadrado de independencia. (14)

Índice de concordancia: en el modelo de ecuación estructural, es una medida de la calidad con que el patrón de correlaciones de una muestra coincide con las correlaciones que se esperarían según el patrón hipotético de causas y efectos entre las variables; usualmente presenta valores de 0 a 1, siendo 1 una concordancia perfecta. (17)

Interpretación de la probabilidad como la frecuencia relativa a largo plazo: comprensión de la probabilidad como la proporción de un determinado resultado que se obtendría si se repitiera el experimento muchas veces. (5)

Interpretación subjetiva de probabilidad: es un modo de entender la probabilidad como el grado de certidumbre de que ocurrirá determinado resultado. (5)

Intervalo de confianza: rango de registros (es decir, los registros que se encuentran entre determinados valores superior e inferior) que probablemente incluye la verdadera media poblacional. (7)

Intervalo de límite abierto: en una tabla de frecuencias agrupadas, es el mayor (o menor) intervalo, que incluye todos los valores por encima (o por debajo) de un determinado valor. (1)

Intervalo de una tabla de frecuencias agrupadas: la serie de valores que se agrupan (por ejemplo, si el tamaño del intervalo fuera 10, uno de los intervalos podría ser de 10,00 a 19,99). (1)

Intervalo del 95% de confianza: intervalo de confianza en el cual, en términos generales, existe un 95% de posibilidades de que se encuentre la media poblacional. (7)

Intervalo del 99% de confianza: intervalo de confianza en el cual, en términos generales, existe un 99% de posibilidades de que se encuentre la media poblacional. (7)

Límites de confianza: valores superior e inferior del intervalo de confianza. (7)

Línea de regresión: una línea en un gráfico que muestra el valor predicho de la variable dependiente correspondiente a cada valor de la variable independiente. (4)

LISREL: programa para computadoras del modelo de ecuación estructural; a veces utilizado como nombre genérico del propio procedimiento. (17)

MANCOVA (Análisis de covarianza multivariado): análisis de covarianza en el que existe más de una variable dependiente. (17)

MANOVA (Análisis de varianza multivariado): análisis de varianza en el que existe más de una variable dependiente. (17)

Mantener constante: es el hecho de anular la influencia de una variable en la asociación entre las otras variables en la regresión múltiple, la correlación parcial o el análisis de covarianza; es igual a **excluir** o **controlar**. (17)

Matriz de correlación: modo común de informar las correlaciones entre diversas variables en publicaciones científicas; tabla con los nombres de las variables en la parte superior y lateral, y en la que se indican todas las correlaciones entre las variables incluidas (por lo general sólo la mitad del cuadrado resultante se completa, ya sea la parte superior o inferior de la diagonal, y la otra mitad resulta redundante). (3)

Media (M, μ): promedio aritmético de un grupo de registros; la suma de los registros dividida por la cantidad de registros. (2)

Media armónica (N'): promedio especial que es influido en mayor medida por los números menores; en una prueba t para medias independientes, cuando las cantidades de registros de los dos grupos son diferentes, se utiliza la media armónica como equivalente del tamaño muestral de cada grupo para la determinación la potencia. (10)

Media de casilla: media de una combinación en particular de niveles de las variables independientes en un diseño factorial. (13)

Media de columna: registro medio de todos los participantes de un nivel determinado de la variable independiente, cuyos niveles forman las columnas en el esquema diagramado de un diseño factorial. (13)

Media de fila: en un diseño factorial es el registro medio de todos los participantes que forman un determinado nivel de la variable independiente, cuyos niveles corresponden a las filas de la disposición diagramada y a niveles que corresponden a las filas en el esquema diagramado. (13)

Media marginal: en un diseño factorial, el registro medio de todos los participantes de un determinado nivel de una de las variables independientes. (13)

Media poblacional (μ): media de la población (usualmente desconocida). (5)

Mediana: el registro medio después de ordenar de mayor a menor todos los registros de una distribución. (2)

Medición de intervalos iguales: medición en la que la diferencia entre cualquier par de valores representa la misma diferencia del aspecto implícito bajo medición. (15)

Medición de rango y orden: medición en la que los valores mayores representan más del aspecto implícito que se está midiendo, pero en la que la diferencia entre dos registros cualesquiera no representa el mismo nivel de diferencia del aspecto implícito que se está midiendo; es igual a la **variable ordinal**. (1, 15)

Medición ordinal: es igual a la **medición de rango y orden**. (1,15)

Meta-análisis: método estadístico para combinar los resultados de estudios independientes, usualmente enfocado en los tamaños de efecto. (8)

Métodos intensivos por computadora: métodos estadísticos que incluyen procedimientos de prueba de hipótesis, los cuales involucran grandes cantidades de cálculos repetidos. Estos métodos sólo se han hecho posibles últimamente gracias a la disponibilidad de computadoras.

Moda: el valor con mayor frecuencia en una distribución. (2)

Modelo causal: en el modelo de ecuación estructural, es la serie de senderos causales entre variables latentes. (17)

Modelo de cuadrados mínimos: método usual de determinación de los valores óptimos de los coeficientes de correlación; esos valores óptimos son los que producen el menor error cuadrático en los valores predichos. (16)

Modelo de ecuación estructural: versión sofisticada del análisis de senderos, que incluye senderos que involucran variables teóricas –no medidas–, latentes, y que además permite realizar una especie de prueba de significación y proporciona medidas de la concordancia general de los datos con el patrón causal hipotético; también se denomina **modelo de variable latente** y **LISREL**. (17)

Modelo de medición: en el modelo de ecuación estructural, es la serie de senderos causales entre la variable latente y la variable manifiesta. (17)

Modelo de predicción: fórmula para realizar predicciones; es decir, fórmula para predecir el registro de una persona en una variable dependiente sobre la base del registro de esa persona en una o más variables independientes. (4)

Modelo de variable latente: es igual al **modelo de ecuación estructural**. (17

Modelo estructural: forma de interpretación del análisis de varianza como una división del desvío de cada registro con respecto a la media general en distintas partes que corresponden a la variación intragrupal (el desvío del registro con respecto a la media de su grupo) e intergrupal (el desvío de la media del grupo al que pertenece el registro con respecto a la media general); *interpretación alternativa* (pero *matemáticamente* equivalente) del análisis de varianza. (12, 13)

Modelo lineal general: fórmula general que es la base de la mayoría de los métodos estadísticos tratados en este texto. La fórmula describe un registro como la suma de una constante, la influencia ponderada de diversas variables y el error; la fórmula es similar a una ecuación de regresión múltiple excepto por el hecho de que incluye el error y porque la suma de las influencias es igual al registro real en la variable dependiente (y no el predicho). (16)

Muestra: los registros de determinado grupo de personas analizadas; generalmente considerado representativo de los registros de alguna población más amplia. (5)

Multicolinealidad: situación que se produce en la regresión múltiple, cuando las variables de predicción están correlacionadas entre sí. (4)

Nivel de significación (α): probabilidad de obtener significación estadística si la hipótesis nula en efecto es verdadera; probabilidad de cometer el error Tipo I. (6–8)

Niveles convencionales de significación ($p < 0{,}05$, $p < 0{,}01$): los niveles de significación (niveles alfa) comúnmente utilizados en psicología. (6)

Niveles de medición: distintos tipos de información numérica implícita provista por una medida, como por ejemplo, de intervalos iguales, de rango y orden o nominal (categórica). (1,15)

Normas: parámetros poblacionales conocidos de pruebas estandarizadas (como una prueba de personalidad o aptitud) que sirven como patrones de comparación para cualquier individuo que realice la prueba. (7)

Parámetro poblacional: valor real de la media, el desvío estándar, etc., de la población (comúnmente los parámetros poblacionales se desconocen, aunque a veces se estiman); los parámetros poblacionales por lo general se simbolizan con letras griegas. (5)

Pendiente: *la inclinación del ángulo* de una línea en un gráfico de dos variables, tal como la línea de regresión en un gráfico de la relación entre una variable dependiente y otra independiente; es la cantidad de unidades que la línea se eleva por cada unidad que cruza (en la regresión con puntuaciones originales, la pendiente es igual a = b). (4)

phi de Cramer: medida de relación entre dos variables nominales; medida del tamaño de efecto para una prueba chi-cuadrado de independencia con una tabla de contingencia mayor a 2 x 2; también se lo conoce como V de Cramer, y a veces se representa en forma escrita como ϕ_C ó V_C. (14)

Plegarse sobre factores: procedimiento del análisis factorial de varianza en el que se ignora una de las dimensiones (variables independientes), reduciendo el análisis total en una dimensión pero manteniendo la misma cantidad total de participantes. (13)

Población: el grupo completo de personas al cual un investigador se propone aplicar los resultados de un estudio; aquel grupo más amplio sobre el cual se realizan inferencias sobre la base de una determinada serie de personas analizadas. (5)

Polígono de frecuencias: gráfico de líneas de una distribución en el que los valores se marcan sobre el eje horizontal y la altura de cada punto corresponde a la frecuencia de ese valor, las líneas comienzan y terminan en el eje horizontal y el gráfico se asemeja al contorno de montañas en el horizonte. (1)

Potencia estadística: probabilidad de que el estudio arroje un resultado significativo si la hipótesis de investigación es verdadera. (8)

Potencia: es igual a la **potencia estadística**. (8)

Predicción bivariada: predicción de registros en una variable sobre la base de registros en otra variable. (4)

Probabilidad (p): la frecuencia relativa con que se espera determinado resultado; la proporción de resultados exitosos en relación con todos los resultados. (5)

Procedimiento Bonferroni: procedimiento de comparación múltiple en el cual el porcentaje total de alfa se divide entre la serie de comparaciones, de modo tal que cada una se prueba con un nivel más exigente de significación. (12)

Producto cruzado de puntuaciones Z: es el resultado de multiplicar la puntuación Z de determinada persona, en una variable, por la puntuación Z de esa misma persona en otra variable; con respecto a un grupo de individuos, el promedio de los productos cruzados de puntuaciones Z entre dos variables es el coeficiente de correlación de esas dos variables. (3)

Promedio ponderado: promedio en el que los registros promediados no tienen la misma influencia sobre el total. (10)

Proporción de varianza justificada (r^2, R^2): un indicador del tamaño del efecto en un análisis de varianza; es lo mismo que **la reducción proporcional de error** en la regresión múltiple. (4, 12, 13, 16)

Prueba chi-cuadrado de independencia: procedimiento de prueba de hipótesis que examina si la distribución de frecuencias de las categorías de una variable nominal no está relacionada con la distribución de frecuencias de las categorías de otra variable nominal. (14)

Prueba chi-cuadrado de la bondad de ajuste: procedimiento de prueba de hipótesis que examina en qué medida una distribución de frecuencias observadas de una variable nominal se ajusta a algún patrón esperado de frecuencias. (14)

Prueba de aleatorización aproximada: alternativa de la prueba de aleatorización para aquellos casos en los que la muestra es demasiado grande como para realizar una prueba de aleatorización que tenga en cuenta cada posible reorganización de los datos obtenidos de la muestra; el método de aleatorización aproximada, intensivo por computadora, genera una gran cantidad de las posibles reorganizaciones de datos; por ejemplo 1.000. (15)

Prueba de aleatorización: procedimiento de prueba de hipótesis (por lo general, se trata de un método intensivo por computadora) que tiene en cuenta cada reorganización posible de los datos de la muestra para determinar si la organización de los datos reales de la muestra podría ocurrir por casualidad. (15)

Prueba de dos colas: el procedimiento de prueba de hipótesis para una hipótesis no **direccional**; situación en la cual el sector de la distribución comparativa en el que se rechazaría la hipótesis nula está dividido entre los dos lados (colas) de la distribución. (6)

Prueba de hipótesis: procedimiento utilizado para determinar si los resultados de un experimento (que analiza una muestra) sustentan determinada teoría o innovación práctica (que se considera aplicable a una población). (6)

Prueba de rango y orden: procedimiento de prueba de hipótesis que utiliza datos ordenados por rangos. (15)

Prueba de una cola: el procedimiento de prueba de hipótesis para una hipótesis direccional; situación en la cual el sector de la distribución comparativa en el que se rechazaría la hipótesis nula se encuentra enteramente en un lado (cola) de la distribución. (6)

Prueba libre de distribución: procedimiento de prueba de hipótesis en la que no existen supuestos en cuanto a la forma de las poblaciones implícitas; es similar a una prueba no paramétrica. (15)

Prueba no paramétrica: procedimiento de prueba de hipótesis que no asume supuestos con respecto a parámetros poblacionales; es similar a una **prueba libre de distribución**. (15)

Prueba paramétrica: procedimiento de prueba de hipótesis ordinario, tal como una prueba *t* o un análisis de varianza, que asume supuestos acerca de la forma y otros parámetros de las poblaciones. (15)

Prueba *t* para medias dependientes: procedimiento de prueba de hipótesis en el que cada participante tiene dos registros (o los participantes forman parejas equiparadas) y se desconoce la varianza poblacional; el procedimiento determina la significación de una hipótesis utilizando registros diferenciales de un sólo grupo de participantes. (9)

Prueba *t* para medias independientes: procedimiento de prueba de hipótesis en el que se prueban dos grupos distintos de personas y en el que se desconoce la varianza poblacional. (10)

Prueba *t* para una sola muestra: procedimiento de prueba de hipótesis en el que se compara una media muestral con una media poblacional conocida, y se desconoce la varianza poblacional. (9)

Prueba *t*: procedimiento de prueba de hipótesis en el que se desconoce la varianza poblacional; compara puntuaciones *t* de una muestra con una distribución comparativa denominada distribución *t*. (9, 10)

Prueba Z: procedimiento de prueba de hipótesis en el cual hay una sola muestra y se conoce la varianza poblacional. (7) **Puntuación estándar:** una puntuación *Z* de una distribución que sigue una curva normal; a veces utilizada para referirse a cualquier puntuación Z. (2)

Punto medio: la mitad de un intervalo en un histograma o polígono de frecuencias basados en una tabla de frecuencias agrupadas; punto que se encuentra justo en el medio entre el comienzo del intervalo y el comienzo del intervalo siguiente. (1)

Punto *t*: en una distribución *t*, es la cantidad de desvíos estándar con respecto a la media; es similar a una puntuación Z. (9)

Puntuación original: Una medición ordinaria (o cualquier otro número de una distribución antes de ser convertido en una puntuación *Z* o transformado de algún otro modo). (2)

Puntuaciones Z: Cantidad de desvíos estándar por encima (o por debajo, si es negativo) de la media de su distribución a la que se encuentra un registro; registro ordinario transformado de forma tal que describe más adecuadamente su ubicación en una distribución. (2)

Razón *F*: en el análisis de varianza, es la razón entre la estimación intergrupal de varianza poblacional y la estimación intragrupal de varianza poblacional; es un registro en la distribución comparativa (una distribución *F*) de un análisis de varianza; también se lo llama simplemente *F*. (11–13)

Reducción proporcional de error (r^2, R^2): es la medida de asociación entre variables que se utiliza cuando se comparan asociaciones obtenidas en diferentes estudios o con diferentes variables; el coeficiente de correlación elevado al cuadrado; es el error cuadrático que se reduce utilizando una norma de predicción para regresión múltiple o bivariada con respecto al error cuadrático; utilizando la media para predecir, expresado como una proporción del error cuadrático al utilizar la media para predecir. Es igual a la **proporción de varianza justificada.** (3, 4, 12, 13, 16, 17)

Registro diferencial: es la diferencia entre el registro de un participante en una prueba y el registro de la misma persona en otra prueba. Por lo general, se trata de un registro posterior menos un registro anterior, en cuyo caso también se lo denomina **registro de cambio**. (9)

Registro muestral de corte: punto en la distribución comparativa según el cual, si es igualado o superado por el registro muestral, se rechazará la hipótesis nula. (6)

Registro: valor correspondiente a un determinado participante con respecto a una variable. (1)

Reglas del tamaño de efecto: reglas acerca de lo que se debe considerar con respecto a un tamaño de efecto pequeño, mediano y grande, basadas en lo que resulta típico de la investigación psicológica; también se conocen como **reglas de Cohen**. (8)

Regresión bivariada: ídem predicción bivariada. (4)

Regresión múltiple gradual: procedimiento exploratorio en el que se prueban todas las potenciales variables de predicción que han sido medidas para descubrir la variable que produce la mejor predicción; luego, cada una de las variables restantes se prueba para descubrir la variable que, en combinación con la primera, produce la mejor predicción; el proceso continúa hasta el momento en que, agregar la mejor variable restante, no produce una mejora significativa. (17)

Regresión múltiple jerárquica: método de regresión múltiple en el cual las variables de predicción se agregan, una o unas pocas por vez, de forma secuencial planificada, permitiendo al investigador calcular la contribución de cada variable sucesiva a la predicción por encima de aquellas ya incluidas. (17)

Regresión múltiple: predicción de registros en una variable (la variable dependiente) sobre la base de registros en otras dos o más variables (variables de predicción o independientes). (4)

Restricción del rango: situación en la cual se calcula una correlación incluyendo en el grupo estudiado sólo una serie limitada de los posibles valores de una de las variables (3)

Resultado: *término* utilizado al discutir la probabilidad, el cual se refiere a la consecuencia de un experimento (o virtualmente cualquier hecho, como por ejemplo, que una moneda caiga cara hacia arriba o que llueva mañana). (5)

Robustez: la medida en la cual determinado procedimiento de prueba de hipótesis es razonablemente preciso aun cuando no se cumplan los supuestos del mismo. (9)

Selección aleatoria: es un método de selección de muestra que utiliza verdaderos procedimientos de azar (lo que generalmente implica que cada persona de la población tiene las mismas posibilidades de ser seleccionada); uno de los métodos es que el investigador comience con una lista completa de todas las personas que forman la población y seleccione un grupo de ellas para ser analizados, utilizando una tabla de números aleatorios; no debe confundirse con la selección casual. (5)

Selección casual: procedimiento de selección de una muestra de individuos para analizar tomando a aquellos que están disponibles o resultan ser los primeros en una lista; no debe confundirse con la selección aleatoria. (5)

Suma de desvíos cuadráticos intergrupales (SS_{Entre}): suma de los desvíos cuadráticos de la media del grupo al que pertenece cada registro con respecto a la media general; es igual a la **suma intergrupal de cuadrados**. (12)

Suma de desvíos cuadráticos intragrupales (SS_{Dentro}): suma de los desvíos cuadráticos de cada registro con respecto a la media de su grupo; es igual a la suma intragrupal de cuadrados. (12)

Suma de desvíos cuadráticos totales (SS_{Total}): en un análisis de varianza, la suma de desvíos cuadráticos de cada registro con respecto a la media general de todos los registros, ignorando por completo el grupo en el que se encuentra el registro. (12) En la regresión, es la suma de las diferencias cuadráticas entre cada registro y el registro predicho cuando se predice utilizando la media. (4)

Suma de errores cuadráticos (SS_{Error}): suma de las diferencias cuadráticas entre cada registro y el registro predicho correspondiente. (4)

Suma intergrupal de cuadrados (SS_{Entre}). Es igual a la **suma de desvíos cuadráticos intergrupales**. (12)

Suma intragrupal de cuadrados (SS_{Dentro}): es igual a la suma de **desvíos cuadráticos intragrupales**. (12)

Suma total de cuadrados (SS_{Total}): suma de las diferencias cuadráticas entre cada registro y la media general de todos los registros; es igual a la **suma de los desvíos cuadráticos de la media** (SS). (4, 12, 13)

Supuesto: condición necesaria para realizar un determinado procedimiento de prueba de hipótesis, tal como el hecho de que una población tenga distribución normal; parte del fundamento matemático para la exactitud de las tablas utilizadas en el proceso de determinación de los valores de corte. (9–15)

Tabla chi-cuadrado: tabla que proporciona los registros de corte en la distribución chi-cuadrado según distintos grados de libertad y niveles de significación. (14)

Tabla de análisis de varianza: cuadro que muestra los principales elementos del cálculo de un análisis de varianza con el método del modelo estructural. (12, 13)

Tabla de áreas de la curva normal: tabla que muestra los porcentajes de registros de una distribución normalmente distribuida, que se ubican entre la media y diversas cantidades de desvíos estándar por encima de ella (puntuaciones Z). (5)

Tabla de contingencia: cuadro bidimensional que muestra las frecuencias en cada combinación de categorías de dos variables nominales. (14)

Tabla de frecuencias agrupadas: tabla de frecuencias en la que se indica la cantidad de participantes para cada intervalo de valores. (1)

Tabla de frecuencias: lista de la cantidad de individuos que presentan cada uno de los diversos valores de determinada variable. (1)

Tabla de potencia: tabla utilizada para el procedimiento de prueba de hipótesis, que indica la potencia estadística de un estudio según los distintos tamaños de efecto y los niveles de significación. (8)

Tabla F**:** tabla que proporciona los registros de corte en la distribución F según distintos grados de libertad y niveles de significación. (11)

Tabla t**:** tabla que indica los registros de corte en la distribución t según distintos grados de libertad y niveles de significación, y según se trate de pruebas de una o dos colas. (9)

Tamaño de efecto de estudios que involucran una o dos medias (d): Cantidad de desvíos estándar poblacionales en los que difieren las medias poblacionales. (8–10)

Tamaño de efecto en el análisis de varianza (f): el desvío estándar de las medias grupales dividido por el desvío estándar de los valores individuales. (11)

Tamaño de efecto: separación (ausencia de superposición) entre poblaciones debido a la variable independiente; aumenta con el aumento de la diferencia entre las medias y disminuye con el aumento del desvío estándar poblacional, pero no se ve afectado por el tamaño de la muestra. (8)

Tamaño de intervalo: en una tabla de frecuencias agrupadas, la diferencia entre el comienzo de un intervalo y el comienzo del siguiente. (1)

Técnica estadística multivariada: procedimiento estadístico que incluye más de una variable dependiente. (17)

Tendencia central: el valor típico o más representativo de un grupo de registros. (2)

Teorema del límite central: principio matemático que establece que la distribución de las sumas (o medias) de registros tomados al azar de cualquier distribución de individuos tendrá tendencia a formar una curva normal. (5,7)

Transformación a la raíz cuadrada: transformación de datos en la que el investigador utiliza la raíz cuadrada de cada registro. (15)

Transformación de datos: aplicación de uno de varios procedimientos matemáticos (tal como calcular la raíz cuadrada) a cada uno de los registros de una muestra; usualmente se realiza para que la distribución muestral se acerque a la normal. (15)

Transformación de rango y orden: convertir una serie de registros en rangos de modo tal que el registro más alto es el rango 1, el siguiente más alto es el rango 2, y así sucesivamente. (15)

Transformación inversa: transformación de datos en la cual el investigador utiliza el número inverso de cada registro (1 dividido por el registro). (15)

Transformación *log*: transformación de datos en la cual el investigador utiliza el logaritmo de cada registro. (15)

Valor: *número o categoría posible que puede presentar un registro.* (1)

Variable categórica: es igual a la **variable nominal**. (1,14)

Variable de predicción (usualmente X): en la regresión múltiple, es la variable que se utiliza para predecir los registros de individuos en otra variable; a veces se la llama **variable independiente**. (4)

Variable dependiente (usualmente Y): variable considerada un efecto; también se utiliza en la regresión para definir toda variable con respecto a la cual se realiza la predicción.(3, 4)

Variable endógena: variable en un análisis de senderos (que incluye un modelo de ecuación estructural) a la cual se dirigen las flechas. (17)

Variable exógena: variable en un análisis de senderos (que incluye un modelo de ecuación estructural) en la cual comienza una cadena causal y a la que no se dirigen flechas dentro del diagrama de senderos. (17)

Variable independiente (usualmente X): variable considerada una causa; además, a veces, en la regresión cualquier variable de predicción sea o no considerada una causa. (3, 4)

Variable latente: en el modelo de ecuación estructural, es una variable teórica no medida que se presume causa implícita de diversas variables efectivamente medidas en el estudio. (17)

Variable manifiesta: en el modelo de ecuación estructural, es la variable ordinaria medida (en contraposición con la variable latente). (17)

Variable nominal: variable cuyos valores son categorías, sin relación numérica (es decir, son nombres en lugar de números); es igual a una **variable categórica**. (1,14)

Variable: una característica determinada que puede presentar distintos valores. (1)

Variables cruzadas: en un diseño factorial, es la situación en la que cada nivel de una variable independiente se mide a cada nivel de la otra variable independiente. (13)

Varianza (SD^2, S^2, σ^2, CM): medida del grado de dispersión de una serie de registros; promedio de los desvíos cuadráticos con respecto a la media. (2, 5, 9, 11)

Varianza de una distribución de diferencias entre medias ($S^2_{\text{Diferencia}}$): es uno de los cálculos que forman parte de la prueba t para medias independientes; es igual a la suma de las varianzas de las distribuciones de medias de cada una de dos muestras. (10)

Varianza de una distribución de medias (S^2_M, σ^2_M): varianza poblacional dividida por la cantidad de casos en cada muestra. (7, 9)

Varianza poblacional (σ^2): varianza de la población (usualmente desconocida). (5)

Glosario de Símbolos

α: nivel de significación; probabilidad de error Tipo I. (8)

β: coeficiente de regresión estandarizado (4); también probabilidad de error Tipo II en la prueba de hipótesis. (8)

μ: media poblacional. (5)

μ_M: media de una distribución de medias. (7)

σ: desvío estándar poblacional. (5)

σ_M: desvío estándar de una distribución de medias. (7)

σ^2: varianza poblacional. (5)

σ^2_M: varianza de una distribución de medias. (7)

Σ: suma de sumar todos los registros que siguen. (2)

φ: coeficiente phi; tamaño de efecto en el análisis chi-cuadrado con tabla de contingencia 2 x 2. (14)

χ^2: dato estadístico chi-cuadrado. (14)

a: constante de regresión. (4)

b: coeficiente de regresión para puntuaciones originales. (4)

d: tamaño de efecto en estudios que incluyen una o dos medias. (8–10)

gl: grados de libertad. (9–14)

gl_1, gl_2, etc.: grados de libertad del primer grupo, del segundo grupo, etc. (10–13)

gl_{Entre}: grados de libertad del numerador en el análisis de varianza. (11)

gl_{Columnas}, gl_{Filas}, $gl_{\text{Interacción}}$: grados de libertad de columnas, filas e interacciones (en el análisis factorial de varianza). (13)

gl_{Total}: grados de libertad totales de todos los grupos. (10–13)

gl_{Dentro}: grados de libertad del denominador en el análisis de varianza. (11)

f: medida del tamaño de efecto en el análisis de varianza. (11)

Razón F: razón entre la estimación intergrupal de varianza poblacional y la estimación intragrupal de varianza poblacional en el análisis de varianza. (11)

GM: media de todos los registros en el análisis de varianza. (11–13)

M: media. (2)

M_1, M_2, etc.: media del primer grupo, del segundo grupo, etc. (10–13)

M_{Columna}, M_{Fila}: media de los registros en determinada columna o determinada fila (en el análisis factorial de varianza). (13)

CM_{Entre}: cuadrados medios intergrupales. (11)

CM_{Columnas}, CM_{Filas}, $CM_{\text{Interacción}}$: cuadrados medios intergrupales de columnas, filas e interacciones. (13)

CM_{Error}: error de los cuadrados medios. (11)

CM_{Dentro}: cuadrados medios intragrupales. (11)

n: cantidad de registros en cada grupo del análisis de varianza. (11)

N: cantidad total de registros. (2)

N_1, N_2, etc.: cantidad de casos en el primer grupo, en el segundo grupo, etc. (10–13)

N': media armónica de dos tamaños de muestras desiguales. (10)

N_{Columnas}, N_{Filas}: cantidad de columnas, cantidad de filas (en el análisis factorial de varianza). (13)

N_{Casillas}: cantidad de casillas en un diseño factorial. (13)

N_{Grupos}: cantidad de grupos en el análisis de varianza.

p: probabilidad. (5)

r: coeficiente de correlación. (3)

r^2: reducción proporcional de error (proporción de varianza justificada) en una regresión bivariada. (3)

R: coeficiente de correlación múltiple. (4, 12)

R^2: reducción proporcional de error (proporción de varianza justificada) en análisis de varianza y regresión múltiple. (4, 12, 13)

R^2_{Columnas}, R^2_{Filas}, $R^2_{\text{Interacción}}$: proporción de varianza justificada (una medida del tamaño del efecto en el análisis factorial de varianza) por las columnas, las filas y la interacción. (13)

S: estimación no sesgada del desvío estándar poblacional. (9)

S^2: estimación no sesgada de la varianza poblacional. (9)

S^2_1, S^2_2, etc.: estimación no sesgada de la varianza poblacional, basada en los registros de la primer muestra, de la segunda muestra, etc. (10–13)

S^2_{Entre}: estimación intergrupal de la varianza poblacional. (11)

S^2_{Columnas}, S^2_{Filas}, $S^2_{\text{Interacción}}$: varianza poblacional estimada intergrupal de columnas, filas, interacción (en el análisis factorial de varianza). (13)

$S_{\text{Diferencia}}$: desvío estándar de la distribución de diferencias entre medias. (10)

$S^2_{\text{Diferencia}}$: varianza de la distribución de diferencias entre medias. (10)

S^2_{Error}: varianza del error. (4, 11)

S_M: desvío estándar de la distribución de medias basado en una varianza poblacional estimada. (9)

S^2_M: varianza de una distribución de medias basada en una varianza poblacional estimada, en el caso de una prueba t; o estimada a partir de la variación entre medias grupales, en el caso de un análisis de varianza. (9, 11)

S^2_{M1}, S^2_{M2}, etc.: varianza de la distribución de medias basada en una estimación combinada de la varianza poblacional, correspondiente a la primera muestra, la segunda muestra, etc. (10, 11)

$S_{\text{Combinada}}$: estimación combinada del desvío estándar poblacional. (10)

$S^2_{\text{Combinada}}$: estimación combinada de la varianza poblacional. (10)

S^2_{Dentro}: estimación intragrupal de la varianza poblacional. (11)

SD: desvío estándar

SD^2: varianza. (2)

SS: suma de desvíos cuadráticos. (2)

SS_{Entre}: suma de desvíos cuadráticos intergrupales. (12)

SS_{Columnas}, SS_{Filas}, $SS_{\text{Interacción}}$: suma de desvíos cuadráticos entre columnas o filas o por efecto de la interacción (en el análisis factorial de varianza). (13)

SS_{Total}: suma total de desvíos cuadráticos con respecto a la media (o con respecto a la gran media, en el análisis de varianza). (4, 12, 13)

SS_{Dentro}: suma de desvíos cuadráticos intragrupales (o dentro de las casillas). (12, 13)

Puntuación t: cantidad de desvíos cuadráticos con respecto a la media en una distribución t. (9)

X: registro en una variable determinada; en la regresión X, es el nombre usual de la variable de predicción o independiente. (1–4)

X_1, X_2, etc.: primera variable independiente o de predicción, segunda variable independiente o de predicción, etc. (4)

$\bar{X}$: media de la variable denominada X. (2)

Y: por lo general, la variable dependiente en una regresión. (3, 4)

$\hat{Y}$: valor predicho de la variable denominada Y. (4)

Z: cantidad de desvíos estándar de la media. (2)

Z_X: puntuación Z de la variable denominada X. (3, 4)

Z_{X1}, Z_{X2}, etc.: puntuación Z de la primera variable independiente o de predicción; puntuación Z de la segunda variable independiente o de predicción, etc. (4)

Z_Y: puntuación Z de la variable denominada Y. (3, 4)

$\hat{Z}_Y$: valor predicho de la puntuación Z en la variable denominada y. (4)

Otros símbolos

$\hat{}$: valor predicho de la variable. (4)

$\bar{}$: media de la variable. (2)

Referencias bibliográficas

ABELSON, R. P. (1997). On the surprising longevity of flogged horses: Why there is a case for the significance test. *Psychological Science*, *8*, 12–15.

ALEXANDER, C. N., LANGER, E. J., NEWMAN, R. I., CHANDLER, H. M., & Davies, J. L. (1989). Transcendental Meditation, mindfulness, and longevity: An experimental study with the elderly. *Journal of Personality and Social Psychology*, *57*, 950–964.

ALGINA, J., & KESELMAN, H. J. (1997). Detecting repeated measures effects with univariate and multivariate statistics. *Psychological Methods*, *2*, 208–218.

ALTMAN, D. G., LEVINE, D. W., HOWARD, G., & HAMILTON, H. (1997). Tobacco farming and public health: Attitudes of the general public and farmers. *Journal of Social Issues*, *53*, 113–128.

AMERICAN PSYCHOLOGICAL ASSOCIATION (1994). *Graduate study in psychology*. Washington, DC: Author.

ARON, A., & ARON, E. N. (1989). *The heart of social psychology*. Lexington, MA: Heath.

ARON, A., & Aron, E. N. (1997). Self-expansion motivation and including other in the self. En W. Ickes (Section Ed.) & S. Duck (Ed.), *Handbook of personal relationships* (2nd Edition, Vol. 1, pp. 251–270). London: Wiley.

ARON, A., ARON, E. N., & ALLEN, J. (in press). Motivations for unreciprocated love. *Personality and Social Psychology Bulletin*.

ARON, A., & FRALEY, B. (in press). Relationship closeness as including other in the self: Cognitive underpinnings and measures. *Social Cognition*.

ARON, A., MELINAT, E., ARON, E. N., VALLONE, R. D., & BATOR, R. J. (1997). The experimental generation of interpersonal closeness: A procedure and some preliminary findings. *Personality and Social Psychology Bulletin*, *23*, 363–377.

ARON, A., PARIS, M., & ARON, E. N. (1995). Falling in love: Prospective studies of self-concept change. *Journal of Personality and Social Psychology*, *69*, 1102–1112.

ARON, E. N. (1996). *The highly sensitive person*. New York: Birch/Lane.

ARON, E. N., & ARON, A. (1997). Sensory processing sensitivity and its relation to introversion and emotionality. *Journal of Personality and Social Psychology*, *73*, 345–368.

BAKER, D. P., & JONES, D. P. (1993). Creating gender equality: Cross-national gender stratification and mathematical performance. *Sociology of Education*, *66*, 91–103.

BARDSLEY, J. J., & RHODES, S. R. (1996). Using the Steers-Rhodes (1984) framework to identify correlates of employee lateness. *Journal of Business and Psychology*, *10*, 351–365.

BAUMRIND, D. (1983). Specious causal attributions in the social sciences: The reformulated stepping-stone theory of heroin use as exemplar. *Journal of Personality and Social Psychology*, *45*, 1289–1298.

BIERNAT, M., & WORTMAN, C. B. (1991). Sharing of home responsibilities between professionally employed women and their husbands. *Journal of Personality and Social Psychology*, *60*, 844–860.

BLANCHARD, F. A., LILLY, T., & VAUGHN, L. A. (1991). Reducing the expression of racial prejudice. *Psychological Science*, *2*, 101–105.

BOYD, C. P., & GULLONE, E. (1997). An investigation of negative affectivity in Australian adolescents. *Journal of Clinical Child Psychology, 26*, 190–197.

BREWER, J. K. (1972). On the power of statistical tests in the *American Education Research Journal. American Educational Research Journal*, 9, 391–401.

BRICKMAN, P., COATES, D., & JANOFF-BULMAN, R. (1978). Lottery winners and accident victims: Is happiness relative? *Journal of Personality and Social Psychology, 36*, 917–927.

BUCK, J. L. (1985). A failure to find gender differences in statistics achievement. *Teaching of Psychology, 12*, 100.

BUSS, D. M., & SCHMITT, D. P. (1993). Sexual strategies theory: An evolutionary perspective on human mating. *Psychological Review, 100*, 204–232.

CAPALDI, D. M., & PATTERSON, G. R. (1991). Relation of parental transitions to boys' adjustment problems: 1. A linear hypothesis 2. Mothers at risk for transitions and unskilled parenting. *Developmental Psychology, 27*, 489–504.

CAREY, M. P., MAISTO, S. A., KALICHMAN, S. C., FORSYTH, A.D., WRIGHT, E. M., & JOHNSON, B. T. (1997). Enhancing motivation to reduce the risk of HIV infection for economically disadvantaged urban women. *Journal of Consulting and Clinical Psychology, 65*, 531–541.

CASPI, A., BEGG, D., DICKSON, N., HARRINGTON, H., LANGLEY, J., MOFFITT, T. E., & SILVA, P. A. (1997). Personality differences predict health-risk behaviors in young adulthood: Evidence from a longitudinal study. *Journal of Personality and Social Psychology, 73*, 1052–1063.

CASPI, A., & HERBENER, E. S. (1990). Continuity and change: Assortative marriage and the consistency of personality in adulthood. *Journal of Personality and Social Psychology, 58*, 250–258.

CATANZARO, D., & TAYLOR, J. C. (1996). The scaling of dispersion and correlation: A comparison of least-squares and absolute deviation statistics. *British Journal of Mathematical and Statistical Psychology, 49*, 171–188.

CHAPMAN, H. A., HOBFOLL, S. E., & RITTER, C. (1997). Partners' stress underestimations lead to women's distress: A study of pregnant inner-city women. *Journal of Personality and Social Psychology, 73*, 418–425.

CHASE, L. J., & CHASE, R. B. (1976). A statistical power analysis of applied psychological research. *Journal of Applied Psychology, 61*, 234–237.

CHIU, C., HONG, Y., & DWECK, C. S. (1997). Lay dispositionism and implicit theories of personality. *Journal of Personality and Social Psychology, 73*, 19–30.

CHOW, S. L. (1988). Significance test or effect size. *Psychological Bulletin, 103*, 105–110.

CHOW, S. L. (1996). *Statistical significance: Rationale, validity, and utility.* London: Sage.

CIALDINI, R. B., BROWN, S. L., LEWIS, B. P., LUCE, C., & NEUBERG, S. L. (1997). Reinterpreting the empathy-altruism relationship: When one into one equals oneness. *Journal of Personality and Social Psychology, 73*, 481–494.

CLARK, D. M., SALKOVSKIS, P. M., OST, L-G., BREITHOLTZ, E., KOEHLER, K. A., WESTLING, B. E., JEAVONS, A., & GELDER, M. (1997). Misinterpretation of body sensations in panic disorder. *Journal of Consulting and Clinical Psychology, 65*, 203–213.

COHEN, J. (1962). The statistical power of abnormal-social psychological research: A review. *Journal of Abnormal and Social Psychology, 65*, 145–153.

COHEN, J. (1988). *Statistical power analysis for the behavioral sciences.* Hillsdale, NJ: Erlbaum.

COHEN, J. (1990). Things I have learned (so far). *American Psychologist, 45*, 1304–1312.

COHEN, J. (1992). A power primer. Psychological *Bulletin, 112*, 155–159.

Cohen, J. (1994). The Earth is round (p < 0,05). American Psychologist, 49, 997–1003.

COHEN, J., & COHEN, P. (1983). *Applied multiple regression/"correlation analysis for the behavioral sciences.* Hillsdale, NJ: Erlbaum.

COMMISSION PAYMENTS TO TRAVEL AGENTS. (1978, August 8). *New York Times*, p. D–1.

CONNORS, G. J., CARROLL, K. M., DICLEMENTE, C. C., LONGABAUGH, R., & DONOVAN, D. M. (1997). The therapeutic alliance and its relationship to alcoholism treatment participation and outcome. *Journal of Consulting and Clinical Psychology*, 65, 588–598.

CONOVER, W., & IMAN, R. L. (1981). Rank transformations as a bridge between parametric and nonparametric statistics. *American Statistician, 35*, 124–129.

COOK, T. D., & CAMPBELL, D. T. (1979). *Quasi-experimentation: Design and analysis issues for field settings.* Skokie, IL: Rand McNally.

COOPER, S. E., & ROBINSON, D. A. G. (1989). The influence of gender and anxiety on mathematics performance. *Journal of College Student Development, 30*, 459–461.

CORTINA, J. M., & DUNLOP, W. P. (1997). On the logic and purpose of significance testing. *Psychological Methods, 2*, 161–172.

CRICK, N. R., CASAS, J. F., & MOSHER, M. (1997). Relational and overt aggression in preschool. *Developmental Psychology, 33*, 579–588.

DAHLSTROM, W. G., LARBAR, D., & DAHLSTROM, L. E. (1986). *MMPI patterns of American minorities*. Minneapolis: University of Minnesota Press.

DANE, F. C., & WRIGHTSMAN, L. S. (1982). Effects of defendants' and victims' characteristics on jurors' verdicts. In N. L. Kerr & R. M. Bray (Eds.), *The psychology of the courtroom*. Orlando, FL: Academic Press.

DARLINGTON, R. B. (1990). *Regression and linear models*. New York: McGraw-Hill.

DAWES, R. M., FAUST, D., & MEEHL, P. E. (1993). Statistical prediction versus clinical prediction: Improving what works. In G. Keren & C. Lewis (Eds.), *A handbook for data analysis in the behavioral sciences: Methodological issues* (pp. 351–367). Hillsdale, NJ: Erlbaum.

DECARLO, L. T. (1997). On the meaning and use of kurtosis. *Psychological Methods, 2*, 292–307.

DEGARMO, D. S., & FORGATCH, M. S. (1997). Determinants of observed confidant support for divorced mothers. *Journal of Personality and Social Psychology, 72*, 336–345.

DELUCCHI, K. L. (1983). The use and misuse of chi-square: Lewis and Burke revisited. *Psychological Bulletin, 94*, 166–176.

DESMARIS, S., & CURTIS, J. (1997). Gender and perceived pay entitlement: Testing for effects of experience with income. *Journal of Personality and Social Psychology, 72*, 141–150.

DUNLAP, W. P., & MYERS, L. (1997). Approximating power for significance tests with one degree of freedom. *Psychological Methods, 2*, 186–191.

DWINELL, P. E., & HIGBEE, J. L. (1991). Affective variables related to mathematics achievement among high-risk college freshmen. *Psychological Reports, 69*, 399–403.

ENDLER, N. S., & MAGNUSSON, D. (1976). Toward an interactional psychology of personality. *Psychological Bulletin, 83*, 956–974.

EPPLEY, K. R., ABRAMS, A. I., & SHEAR, J. (1989). Differential effects of relaxation techniques on trait anxiety: A meta-analysis. *Journal of Clinical Psychology, 45*, 957–974.

ESCUDERO, V., ROGERS, L. E., & GUTIERREZ, E. (1997). Patterns of relational control and nonverbal affect in clinic and nonclinic couples. *Journal of Social and Personal Relationships, 14*, 5–29.

EVANS, R. (1976). *The making of psychology*. New York: Knopf.

EVERETT, S. A., PRICE, J. H., BEDELL, A. W., & TELLJOHANN, S. K. (1997). The effect of a monetary incentive in increasing the return rate of a survey to family physicians. *Evaluation and the Health Professions, 20*, 207–214.

EYSENCK, H. J. (1981). *A model for personality*. Berlin: Springer-Verlag.

FAWZI, M. C. S., PHAM, T., LIN, L., NGUYEN, T. V., NGO, D., MURPHY, E., & MOLLICA, R. F. (1997). The validity of posttraumatic stress disorder among Vietnamese refugees. *Journal of Traumatic Stress, 10*, 101–108.

FISHER, B. (1978). *Fisher Divorce Adjustment Scale*. Boulder, CO: Family Relations Learning Center.

FISHER, R. A. (1938). *Statistical methods for research workers* (7th ed.). London: Oliver & Boyd.

FOERTSCH, J., & GERNSBACHER, M. A. (1997). In search of gender neutrality: Is singular they a cognitively efficient substitute for generic *he*? *Psychological Science, 8*, 106–112.

FOLWELL, A. L., CHUNG, L. C., NUSSBAUM, J. F., BETHEA, L.S., & GRANT, J. A. (1997). Differential accounts of closeness in older adult sibling relationships. *Journal of Social and Personal Relationships, 14*, 843–849.

FORD, J. D., FISHER, P., & LARSON, L. (1997). Object relations as a predictor of treatment outcome with chronic posttraumatic stress disorder. *Journal of Consulting and Clinical Psychology, 65*, 547–559.

FRANK, S. J., POORMAN, M. O., VAN EGEREN, L. A., & FIELD, D. T. (1997). Perceived relationships with parents among adolescent inpatients with depressive preoccupations and depressed mood. *Journal of Clinical Child Psychology, 26*, 205–215.

FRANZ, M. L. VON. (1979). *The problem of puer aeternus*. New York: Springer-Verlag.

FRICK, R. W. (1995). Accepting the null hypothesis. *Memory and Cognition, 23*, 132–138.

FRICK, R. W. (1996). The appropriate use of null hypothesis testing. *Psychological Methods, 1*, 379–390.

FRICK, R. W. (in press). Interpreting statistical testing: Process and propensity, not population and random sampling. *Behavior Research Methods, Instruments, and Computers.*

FRISCH, A. S., SHAMSUDDIN, K., & KURTZ, M. (1995). Family factors and knowledge: Attitudes and efforts concerning exposure to environmental tobacco among Malaysian medical students. *Journal of Asian and African Studies, 30,* 68–79.

GALLUP, D. G. H. (1972). *The Gallup poll: Public opinion, 1935–1971.* New York: Random House.

GALTON, F. (1889). *Natural inheritance.* London: Macmillan.

GAMES, P. A. (1988). Theory-free statistics and theory-based statistics: Their appropriate roles in the reporting of scientific results. *Journal of Experimental Education, 57,* 47–58.

GEISSER, S., & GREENHOUSE, S. (1958). An extension of Box's results on the use of the *F* distribution in multivariate analysis. *Annals of Mathematical Statistics, 29,* 885–891.

GIGERENZER, G., & MURRAY, D. J. (1987). *Cognition as intuitive statistics.* Hillsdale, NJ: Erlbaum.

GIGERENZER, G., SWIJTINK, Z., PORTER, Y., DASTON, L., BEATTY, J., & KRUGER, L. (1989). *The empire of chance.* Cambridge, England: Cambridge University Press.

GIRE, J. T. (1997). The varying effect of individualism-collectivism on preference for methods of conflict resolution. *Canadian Journal of Behavioural Science, 29,* 38–43.

GOSSET, W. S. (1947). *"Student's" collected papers.* London: University College.

GOUGH, H., & HEILBRUN, A. (1983). *The Adjective Check List Manual.* Palo Alto, CA: Consulting Psychologist Press.

GRAHAM, S., WEINER, B., & ZUCKER, G. S. (1997). An attributional analysis of punishment goals and public reactions to O. J. Simpson. *Personality and Social Psychology Bulletin, 23,* 331–346.

GREENWALD, A. G. (1975). Consequences of prejudice against the null hypothesis. *Psychological Bulletin, 82,* 1–19.

GRILO, C. M., WALKER, M. L., BECKER, D. F., EDELL, W. S., & MCGLASHAN, T. H. (1997). Personality disorders in adolescents with major depression, substance use disorders, and coexisting major depression and substance use disorders. *Journal of Consulting and Clinical Psychology, 65,* 328–332.

GUMP, B. B., & KULIK, J. A. (1997). Stress, affiliation, and emotional contagion. *Journal of Personality and Social Psychology, 72,* 305–319.

HAMILTON, D. (1981). *Cognitive processes in stereotyping and intergroup behavior.* Hillsdale, NJ: Erlbaum.

HAMILTON, D., & GIFFORD, R. (1976). Illusory correlation in interpersonal perception: A cognitive basis of stereotypic judgments. *Journal of Experimental Social Psychology, 12,* 392–407.

HARRIS, R. J. (1997). Significance tests have their place. *Psychological Science, 8,* 8–11.

HARTER, S., WATERS, P. L., PETTITT, L. M., WHITESELL, N., KOFKIN, J., & JORDAN, J. (1997). Autonomy and connectedness as dimensions of relationship styles in men and women. *Journal of Social and Personal Relationships, 14,* 147–164.

HAZAN, C., & SHAVER, P. (1987). Romantic love conceptualized as an attachment process. *Journal of Personality and Social Psychology, 52,* 511–524.

HERMANN, C., BLANCHARD, E. B., & FLOR, H. (1997). Biofeedback treatment for pediatric migraine: Prediction of treatment outcome. *Journal of Consulting and Clinical Psychology, 65,* 611–616.

HIGHLEN, P. S., & FINLEY, H. C. (1996). Doing qualitative analysis. En: N. F. T. L. Leong & J. T. Austin (Eds.), *The psychology research handbook* (pp. 177–192). Thousand Oaks, CA: Sage.

HILGARD, E. R. (1987). *Psychology in America: A historical perspective.* Orlando, FL: Harcourt Brace Jovanovich.

HINDLEY, C., FILLIOZAT, A., KLACKENBERG, G., NICOLET-MEISTER, D., & SAND, E. (1966). Differences in age of walking in five European longitudinal samples. *Human Biology, 38,* 364–379.

HOLDEN, G. W., THOMPSON, E. E., ZAMBARANO, R. J., & MARSHALL, L. A. (1997). Child effects as a source of change in maternal attitudes. *Journal of Social and Personal Relationships, 14,* 481–490.

HOLZWORTH, R. J. (1996). Policy capturing with ridge regression. *Organizational Behavior and Human Decision Processes, 68,* 171–179.

HOPKINS, K. D., & GLASS, G. V. (1978). *Basic statistics for the behavioral sciences.* Englewood Cliffs, NJ: Prentice Hall.

HUNTER, J. E. (1997). Needed: A ban on the significance test. *Psychological Science, 8,* 3–7.

HUSSERL, E. (1970). *The crisis of European sciences and transcendental phenomenology: An introduction to phenomenological philosophy* (D. C. Carr, Trans.). Evanston, IL: Northwestern University Press.

HYDE, J. S. (1993). Gender differences in mathematics ability, anxiety, and attitudes: What do meta-analyses tell us? En: L. A. Penner, G. M. Batsche, H. M. Knoff,

& D. L. Nelson, *The challenge in mathematics and science education: Psychology's response* (pp. 237–249). Washington, DC: American Psychological Association.

HYDE, J. S., FENNEMA, E., & LAMON, S. J. (1990). Gender differences in mathematics performance: A meta-analysis. *Psychological Bulletin, 107*, 139–155.

INHOFF, A., LIMA, S., & CARROLL, P. (1984). Contextural effects on metaphor comprehension in reading. *Memory and Cognition, 12*, 558–567.

JEHN, K. A., & SHAH, P. P. (1997). Interpersonal relationships and task performance: An examination of mediating processes in friendship and acquaintance groups. *Journal of Personality and Social Psychology, 72*, 775–790.

JESSOR, R. (1996). Ethnographic methods in contemporary perspective. En: R. Jessor, A. Colby, & R. A. Shweder (Eds.), *Ethnography and human development: Context and meaning in social inquiry* (pp. 3 – 14). Chicago: University of Chicago Press.

JOHNSON, C., & MULLEN, B. (1994). Evidence for the accessibility of paired distinctiveness in distinctiveness-based illusory correlaion in stereotyping. *Personality and Social Psychology Bulletin, 20*, 65–70.

JUDD, C. M., MCCLELLAND, G. H., & CULHANE, S. E. (1995). Data analysis: Continuing issues in the everyday analysis of psychological data. *Annual Review of Psychology, 46*, 433–465.

KAGAN, J. (1994). *Galen's prophecy: Temperament in human nature*. New York: Basic Books.

KARNEY, B. R., & BRADBURY, T. N. (1997). Neuroticism, marital interaction, and the trajectory of marital satisfaction. *Journal of Personality and Social Psychology, 72*, 1075–1092.

KELLEY, H. H. (1971). *Attribution in social interaction*. Morristown, NJ: General Learning Press.

KENNEY, D. A. (1995). Relationship science in the 21st century. *Journal of Social and Personal Relationships, 12*, 597–600.

KERLINGER, F. N. (1973). *Foundations of behavioral research*. New York: Holt, Rinehart and Winston.

KESELMAN, J. C., LIX, L. M., & KESELMAN, H. J. (1996). The analysis of repeated measurements: A quantitative research synthesis. *British Journal of Mathematical and Statistical Psychology, 49*, 275–298.

KLEIJN, W. C., VAN DER PLOEG, H. M., & TOPMAN, R. M. (1994). Cognition, study habits, test anxiety, and academic performance. *Psychological Reports, 75*, 1219–1226.

KLEINMUNTZ, B. (1990). Why we still use our heads instead of formulas: Toward an integrative approach. *Psychological Bulletin, 107*, 296–310.

KRAEMER, H. C., & THIEMANN, S. (1987). *How many subjects? Statistical power analysis in research*. Newbury Park, CA: Sage.

KUNDA, Z., & OLESON, K. C. (1997). When exceptions prove the rule: How extremity of deviance determines the impact of deviant examples on stereotypes. *Journal of Personality and Social Psychology, 72*, 965–979.

KWAN, V. S. Y., BOND, M. H., & SINGELIS, T. M. (1997). Pancultural explanations for life satisfaction: Adding relationship harmony to self-esteem. *Journal of Personality and Social Psychology, 73*, 1038–1051.

LAMBERT, A. J., KHAN, S. R., LICKEL, B. A., & FRICKE, K. (1997). Mood and the correction of positive versus negative stereotypes. *Journal of Personality and Social Psychology, 72*, 1002–1016.

LEVENTHAL, L., & HUYN, C-L. (1996). Directional decisions for two-tailed tests: Power, error rates, and sample size. *Psychological Methods, 1*, 278–292.

LEWIS, D., & BURKE, C. J. (1949). The use and misuse of the chi-square test. *Psychological Bulletin, 46*, 433–489.

LINDZEY, E. W., MIZE, J., & PETTIT, G. S. (1997). Mutuality in parent-child play: Consequences for children's peer competence. *Journal of Social and Personal Relationships, 14*, 523–538.

LYDON, J., PIERCE, T., & O'REGAN, S. (1997). Coping with moral commitment to long-distance dating relationships. *Journal of Personality and Social Psychology, 73*, 104–113.

MACDONALD, C., CHAMBERLAIN, K., & LONG, N. (1997). Race, combat, and PTSD in a community sample of New Zealand Vietnam War veterans. *Journal of Traumatic Stress, 10*, 117–124.

MACKINNON-LEWIS, C., STARNES, R., VOLLING, B., & JOHNSON, S. (1997). Perceptions of parenting as predictors of boys' sibling and peer relations. *Developmental Psychology, 33*, 1024–1031.

MARKMAN, H. J., RENICK, M. J., FLOYD, F. J., STANLEY, S. M., & CLEMENTS, M. (1993). Preventing marital distress through communication and conflict management training: A 4- and 5-year follow-up. *Journal of Consulting and Clinical Psychology, 61*, 70–77.

MAXWELL, S. E., & DELANEY, H. D. (1990). *Designing experiments and analyzing data*. Belmont, CA: Wadsworth.

MAXWELL, S. E., & DELANEY, H. D. (1993). Bivariate median splits and spurious statistical significance. *Psychological Bulletin, 113,* 181–190.

McCONNELL, A. R., SHERMAN, S. J., & HAMILTON, D. L. (1994). Illusory correlation in the perception of groups: An extension of the distinctiveness-based account. *Journal of Personality and Social Psychology, 67,* 414–429.

McCRACKEN, G. (1988). *The long interview.* London: Sage.

McLAUGHLIN-VOLPE, T., ARON, A, & REIS, H. T. (1998, August). Closeness during interethnic social interactions and prejudice: A diary study. En: A. Aron (Chair), *Intergroup contact and personal relationships.* Symposium conducted at the Annual Convention of the American Psychological Association, San Francisco, CA.

McLEOD, J. (1996). Qualitative research methods in counseling psychology. En: R. Woolfe & W. Dryden (Eds.), *Handbook of counseling psychology* (pp. 65–86). London: Sage.

MEEHL, P. E. (1954). *Clinical versus statistical prediction: A theoretical analysis and a review of the evidence.* Minneapolis: University of Minnesota Press.

MICCERI, T. (1989). The unicom, the normal curve, and other improbable creatures. *Psychological Bulletin,* 105, 156–166.

MICKELSON, K. D., KESSLER, R. C., & SHAVER, P. R. (1997). Adult attachment in a nationally representative sample. *Journal of Personality and Social Psychology, 73,* 1092–1106.

MIKULINCER, M. (1998). Attachment working models and the sense of trust: An exploration of interaction goals and affect regulation. *Journal of Personality and Social Psychology, 74.*

MILLER, L. C., & FISHKIN, S. A. (1997). On the dynamics of human bonding and reproductive success: Seeking windows on the adapted-for human-environmental interface. En: J. Simpson & D. T. Kenrick (Eds.), *Evolutionary social psychology* (pp. 197–235). Hillsdale, NJ: Erlbaum.

MILLER, R. S. (1997). Inattentive and contented: Relationship commitment and attention to alternatives. *Journal of Personality and Social Psychology,* 73, 758–766.

MILLIGAN, G. W., WONG, D. S., & THOMPSON, P. A. (1987). Robustness properties of nonorthogonal analysis of variance. *Psychological Bulletin, 101,* 464–470.

MISCHEL, W. (1968). Personality and assessment. New York: Wiley.

MIZE, J., & PETTIT, G. S. (1997). Mothers' social coaching, mother-child relationship style, and children's peer competence: Is the medium the message? *Child Development, 68,* 312–332.

MORIARTY, S. E., & EVERETT, S-L. (1994). Commercial breaks: A viewing behavior study. *Journalism Quarterly, 71,* 346–355.

MUELLER, J. H., ELSER, M. J., & ROLLACK, D. N. (1993). Test anxiety and implicit memory. *Bulletin of the Psychonomic Society, 31,* 531–533.

MYERS, D. G. (1991). Union is strength: A consumer's view of meta-analysis. *Personality and Social Psychology Bulletin, 17,* 265–266.

NEZLEK, J. B., KOWALSKI, R. M., LEARY, M. R., BLEVINS, T., & HOLGATE, S. (1997). Personality moderators of reactions to interpersonal rejection: Depression and trait self-esteem. *Personality and Social Psychology Bulletin, 23,* 1235–1244.

NORCROSS, J. C., HANYCH, J. M., & TERRANOVA, R. D. (1996). Graduate study in psychology: 1992–1993. *American Psychologist, 51,* 631–643.

NORMAN, C., & ARON, A. (1997, June). Shared expansion experiences and relationship satisfaction. En: C. Norman (Chair), *How to make your relationship work: Speculations based on hard research.* Symposium conducted at the International Network on Personal Relationships Conference, Oxford, OH.

OAKES, M. (1982). Intuiting strength of association from a correlation coefficient. *British Journal of Psychology, 73,* 51–56.

OLTHOFF, R. K. (1989). *The effectiveness of premarital communication training.* Unpublished doctoral dissertation, California Graduate School of Family Psychology, San Francisco.

ONWUEGBUZIE, A. J. (1995). Statistics text anxiety and female students. *Psychology of Women Quarterly, 19,* 413–418.

ORBACH, I., MIKULINCER, M., KING, R., COHEN, D., & STEIN, D. (1997). Thresholds and tolerance of physical pain in suicidal and nonsuicidal adolescents. *Journal of Consulting and Clinical Psychology, 65,* 646–652.

PEARSON, K. (1978). *The history of statistics in the 17th and 18th centuries.* London: Griffin.

PETERS, W. S. (1987). *Counting for something: Statistical principles and personalities.* New York: Springer-Verlag.

PEZDEK, K., FINGER, K., & HODGE, D. (1997). Planting false childhood memories: The role of event plausibility. *Psychological Science, 8,* 437–441.

PHILLIPS, L. D. (1973). *Bayesian statistics for social scientists.* London: Nelson.

PRENTICE, D. A., & MILLER, D. T. (1992). When small effects are impressive. *Psychological Bulletin, 112,* 160–164.

REBER, P. J., & KOTOVSKY, K. (1997). Implicit learning in problem solving: The role of working memory capacity. *Journal of Experimental Psychology: General, 126,* 178–203.

REGIER, D., MYERS, J., KRAMER, M., ROBINS, L., BLAZER, D., HOUGH, R., EATON, W., & LOCKE, B. (1984). The NIMH Epidemiologic Catchment Area Program. *Archives of General Psychiatry, 41,* 934–941.

REIS, H. T., & STILLER, J. (1992). Publication trends in JPSP: *A three-decade review. Personality and Social Psychology Bulletin, 18,* 465–472.

REISSMAN, C., ARON, A., & BERGEN, M. R. (1993). Shared activities and marital satisfaction: Causal direction and self-expansion versus boredom. *Journal of Social and Personal Relationships, 10,* 253–254.

RIEHL, R. J. (1994). Academic preparation, aspirations, and first-year perfomarmance of first-generation students. *College and University, 70,* 14–19.

ROSNOW, R. L., & ROSENTHAL, R. (1989a). Definition and interpretation of interaction effects. *Psychological Bulletin, 105,* 143–146.

ROSNOW, R. L., & ROSENTHAL, R. (1989b) Statistical procedures and the justification of knowledge in psychological science. *American Psychologist, 44,* 1276–1284.

ROSS, D. C., & KLEIN, D. F. (1988). Group matching: Is this a research technique to be avoided? *Educational and Psychological Measurement, 48,* 281–295.

RUSSELL, J. A. (1991). In defense of a prototype approach to emotion concepts. *Journal of Personality and Social Psychology, 60,* 37–47.

SANBONMATSU, D. M., POSAVAC, S. S., & STASNEY, R. (1997). The subjective beliefs underlying probability overestimation. *Journal of Experimental Social Psychology, 33,* 276–295.

SAWILOWSKY, S. S., & BLAIR, R. C. (1992). A more realistic look at the robustness and Type II error properties of the *t* test to departures from population normality. *Psychological Bulletin, 111,* 352–360.

SCARR, S. (1997). Rules of evidence: A larger context for the statistical debate. *Psychological Science, 8,* 16–17.

SCHMIDT, F. L. (1996). Statistical significance testing and cumulative knowledge in psychology: Implications for training of researchers. *Psychological Methods, 1,* 115–129.

SCHMIDT, F. L., & HUNTER, J. E. (1997). Eight common but false objections to the discontinuation of significance testing in the analysis of research data. En: L. L. Harlow, S. A. Mulaik, & J. H. Steiger (Eds.), *What if there were no significance tests?* (pp. 37–64). Mahwah, NJ: Erlbaum.

SCHNEIDER, W. J., & NEVID, J. S. (1993). Overcoming math anxiety: A comparison of stress inoculation training and systematic desensitization. *Journal of College Student Development, 34,* 283–287.

SEDLMEIER, P., & GIGERENZER, G. (1989). Do studies of statistical power have an effect on the power of studies? *Psychological Bulletin, 105,* 309–316.

SELLERS, R. M., ROWLEY, S. A. J., CHAVOUS, T. M., SHELTON, J. N., & SMITH, M. A. (1997). Multidimensional Inventory of Black Identity: A preliminary investigation of reliability and construct validity. *Journal of Personality and Social Psychology, 73,* 805–815.

SHAPIRO, D. A., & SHAPIRO, D. (1983). Comparative therapy outcome research: Methodological implications of meta-analysis. *Journal of Consulting and Clinical Psychology, 51,* 42–53.

SHAPIRO, J. P., DORMAN, R. L., BURKEY, W. M., WELKER, C. J., & CLOUGH, J. B. (1997). Development and factor analysis of a measure of youth attitudes toward guns and violence. *Journal of Clinical Child Psychology, 26,* 311–320.

SHI, L., SAMUELS, M. E., RICHTER, D. L., STOSKOPF, C. H., BAKER, S. L., & SY, F. (1997). Primary care physicians and barriers to providing care to persons with HIV/AIDS. *Evaluation and the Health Professions, 20,* 164–187.

SHREIDER, YU. A. (1966). Preface to the English edition. En: N.P. Bushlenko, D. I. Golenko, Yu. A. Shreider, L. M. Sobol, & V. G. Sragovich (Yu. A. Shreider, Ed.), *The Monte Carlo method: The method of statistical trials* (G.J. Tee, Trans.), (p. vii). Elmsford, NY: Pergamon Press.

SIEGEL, M., & BIENER, L. (1997). Evaluating the impact of statewide anti-tobacco campaigns: The Massachusetts and California Tobacco Control Programs. *Journal of Social Issues, 53,* 147–168.

SIGALL, H., & Ostrove, N. (1975). Beautiful but dangerous: Effects of offender attractiveness and nature of the crime on juridic judgments. *Journal of Personality and Social Psychology, 31,* 410–414.

SKINNER, B. F. (1956). A case history in scientific method. *American Psychologist, 11,* 221–233.

SPEED, A., & GANGSTEAD, S. W. (1997). Romantic popularity and mate preferences: A peer-nomination study. *Personality and Social Psychology Bulletin, 23,* 928–937.

STEEN, L. A. (1987). Forward. En: S. Tobias, *Succeed with math: Every student's guide to conquering math anxiety* (pp. xvii–xviii). New York: College Entrance Examination Board.

STEERING COMMITTEE OF THE PHYSICIANS HEALTH STUDY RESEARCH GROUP. (1988). Preliminary report: Findings from the aspirin component of the ongoing Physicians Health Study. *New England Journal of Medicine, 318,* 262–264.

STEIL, J. M., & HAY, J. L. (1997). Social comparison in the work place: A study of 60 dual-career couples. *Personality and Social Psychology Bulletin, 23,* 427–438.

STERNBERG, R. J. (1988). *Construct validation of a triangular theory of love.* Unpublished manuscript, Yale University, Department of Psychology, New Haven, CT.

STIGLER, S. M. (1986). *The history of statistics.* Cambridge, MA: Belknap Press.

STIPEK, D. J., & RYAN, R. H. (1997). Economically disadvantaged preschoolers: Ready to learn but further to go. *Developmental Psychology, 33,* 711–723.

STRAHAN, R. F. (1991). Remarks on the binomial effect size display. *American Psychologist, 46,* 1083–1084.

SUH, E., DIENER, E., & FUJITA, F. (1996). Events and subjective well-being: Only recent events matter. *Journal of Personality and Social Psychology, 70,* 1091–1102.

TABACHNICK, B. G., & FIDELL, L. S. (1996). *Using multivariate statistics* (3rd ed.). New York: Harper & Row.

TANKARD, J., Jr. (1984). *The statistical pioneers.* Cambridge, MA: Schenkman.

TERPSTRA, D. E., & ROZELL, E. J. (1997). Sources of human resource information and the link to organizational profitability. *Journal of Applied Behavioral Science, 33,* 66–83.

THOMPSON, K. N., & SCHUMACKER, R. E. (1997). An evaluation of Rosenthal and Rubin's binomial effect size display. *Journal of Educational and Behavioral Statistics, 22,* 109–117.

TOBIAS, S. (1987). *Succeed with math: Every student's guide to conquering math anxiety.* New York: College Entrance Examination Board.

TUFTE, E. R. (1983). *The visual display of quantitative information.* Cheshire, CT: Graphic Press.

U.S. BUREAU OF THE CENSUS. (1990). *Statistical abstracts of the United States.* Washington, DC: U.S. Government Printing Office.

U.S. DEPARTMENT OF EDUCATION. (1990). *The condition of education.* Washington, DC: U.S. Government Printing Office.

VALENZUELA, M. (1997). Maternal sensitivity in a developing society: The context of urban poverty and infant chronic undernutrition. *Developmental Psychology, 33,* 845–855.

VAN AKEN, M. A. G., & ASENDORPF, J. B. (1997). Support by parents, classmates, friends, and siblings in preadolescence: Covariation and compensation across relationships. *Journal of Social and Personal Relationships, 14,* 79–93.

VAN LANGE, P. M., OTTEN, W., DEBRUIN, E. M. N., & Joireman, J. A. (1997). Development of prosocial, individualistic, and competitive orientations: Theory and preliminary evidence. *Journal of Personality and Social Psychology, 73,* 733–746.

VANCE, W. R., Jr., & WATSON, T. S. (1994). Comparing anxiety management training and systematic rational restructuring for reducing mathematics anxiety in college students. *Journal of College Student Development, 35,* 261–266.

WALBERG, H. J., STRYKOWSKI, B. F., ROVAI, E., & HUNG, S. S. (1984). Exceptional performance. *Review of Educational Research,* 54, 87–112.

WARDLE, J., STEPTOE, A., BELLISLE, F., DAVOU, B., RESCHKE, K., LAPPALAINEN, R., & FREDRIKSON, M. (1997). Healthy dietary practices among European students. *Health Psychology, 16,* 443–450.

WATTS, W., & WRIGHT, L. (1990). The relationship of alcohol, tobacco, marijuana, and other illegal drug use to delinquency among Mexican-American, black, and white adolescent males. *Adolescence, 25,* 171–181.

WECHSLER, H., DAVENPORT, A., DOWDALL, G., MOEYKENS, B., & CASTILLO, S. (1994). Health and behavioral consequences of binge drinking in college: A national survey of students at 140 campuses. *Journal of the American Medical Association, 272,* 1672–1677.

WELLER, A., & WELLER, L. (1997). Menstrual synchrony under optimal conditions: Bedouin families. *Journal of Comparative Psychology, 111,* 143–151.

WHITECOTTON, S. M. (1996). The effects of experience and a decision aid on the slope, scatter, and bias of earnings forecasts. *Organizational Behavior and Human Decision Processes, 66,* 111–121.

WISEMAN, H. (1997). Interpersonal relatedness and self-definition in the experience of loneliness during the transition to university. *Personal Relationships, 4*, 285–299.

WONG, M. M., & CSIKSZENTMIHALYI, M. (1991). Affiliation motivation and daily experience: Some issues on gender differences. *Journal of Personality and Social Psychology, 60*, 154–164.

ZEIDNER, M. (1991). Statistics and mathematics anxiety in social science students: Some interesting parallels. *British Journal of Education, 61*, 319–329.

Índice analítico

B

C

D

E

F

G

H

I

J

K

N

O

P

Q

R

S

T

U

V

W

X

Y

Pearson Educación

Visítenos en www.pearsonedlatino.com

Argentina
Av. Regimiento de los Patricios 1959
(C1266AAF) Buenos Aires
Argentina
Tel. (54-11) 4309-6100
Fax (54-11) 4309-6199
E-mail: info@pearsoned.com.ar

América Central-Panamá
Barrio La Guaria, Moravia
75 metros norte,
Del Portón Norte del Club La Guaria
San José, Costa Rica
Tel. (506) 235 72 76
Fax (506) 297 28 52
E-mail: envwong@racsa.co.cr

Brasil
Rua Emilio Goeldi 747, Lapa
(05065-110) São Paulo - SP
Brasil
Tel. (5511) 36111-0201
Fax (5511) 36111-0654

Caribe
Monte Mall, 2do. piso, suite 21-B
Av. Muñoz Rivera
Hato Rey
Puerto Rico 00918-4261
Tel. (787) 751-4830
Fax (787) 751-1677
E-mail: awlcarib@caribe.net
y awlcarib@caribe.net

Chile
Av Manuel Montt 1452
Providencia
Santiago, Chile
Tel. (562) 269 2089
Fax (562) 274 6158
E-mail: infopear@pearsoned.cl

Colombia
Carrera 68 # 22-55
Santa Fé de Bogotá, D.C.
Colombia
Tel. (571) 405-9300
Fax (571) 405-9330

España
Nuñez de Balboa 120
(28006) Madrid
España
Tel. (3491) 590-3432
Fax (3491) 590-3448

Estados Unidos
One Lake Street
Upper Saddle River
NJ 07458
Tel. (201) 236-7000
Fax: (201) 236-3400

México
Calle Cuatro No. 25 2do piso
Fracc. Industrial Alce Blanco
(53370), Naucalpan de Juárez
Estado de México
Tel. (305) 3870700
Fax (525) 3870811

Uruguay
Casa Juana de América
Av. 8 de Octubre 3061
(11600) Montevideo
Uruguay
Tel./Fax (5982) 486-1617