ResearchGate

See discussions, stats, and author profiles for this publication at: https://www.researchgate.net/publication/286931297

Estadística para las Ciencias Agropecuarias: Sexta Edición

Book · May 2005

CITATIONS
2

READS
9,962

7 authors, including:

Julio Alejandro Di Rienzo
National University of Cordoba, Argentina
212 PUBLICATIONS **2,742** CITATIONS
SEE PROFILE

Fernando Casanoves
CATIE - Centro Agronómico Tropical de Investigación y Enseñanza
334 PUBLICATIONS **6,731** CITATIONS
SEE PROFILE

Margot Tablada
National University of Cordoba, Argentina
51 PUBLICATIONS **1,022** CITATIONS
SEE PROFILE

Maria del Pilar Diaz
National University of Cordoba, Argentina
126 PUBLICATIONS **2,457** CITATIONS
SEE PROFILE

Some of the authors of this publication are also working on these related projects:

Project Climate change mitigation mechanisms, capacity building and technology transfer in Latin America View project

Project Estadística para las Ciencias Agropecuarias View project

All content following this page was uploaded by Mónica Balzarini on 16 December 2015.

The user has requested enhancement of the downloaded file.

Estadística para las Ciencias Agropecuarias

Sexta Edición

Di Rienzo, Julio Alejandro

Casanoves, Fernando

Gonzalez, Laura Alicia

Tablada, Elena Margot

Díaz, María del Pilar

Robledo, Carlos Walter

Balzarini, Mónica Graciela

SEXTA EDICIÓN

Primera Impresión

EDICIÓN ELECTRÓNICA

Julio Di Rienzo

Fernando Casanoves

© by Di Rienzo, Julio Alejandro; Casanoves, Fernando; Gonzalez, Laura Alicia; Tablada, Elena Margot; Díaz, María del Pilar; Robledo, Carlos Walter; Balzarini, Mónica Graciela.

ISBN: 987-1142-68-4

Queda hecho el depósito que prevé la ley 11.723

Queda prohibida la reproducción total o parcial de este libro en forma idéntica o modificada por cualquier medio mecánico o electrónico, incluyendo fotocopia, grabación o cualquier sistema de almacenamiento y recuperación de información no autorizada por los autores.

Prólogo

La estadística aplicada ha tenido un gran florecimiento en los últimos 20 años y hoy es parte del lenguaje científico cotidiano. Aunque el tratamiento estadístico de los resultados experimentales no es un seguro contra los hallazgos casuales, es un gran avance en ese sentido y representa una formidable herramienta para la interpretación de datos, no solo poniendo restricciones a la percepción caprichosa de la información, sino guiando metodológicamente su indagación.

La enseñanza de la estadística en las ciencias agropecuarias no es un tributo a la modernidad sino una larga tradición que se origina en los trabajos de Fisher que, a comienzos del siglo XX, sentaron las bases de la estadística aplicada a la experimentación agrícola.

La sexta edición es el resultado de un trabajo de reorganización de contenidos, selección y actualización de ejemplos y reformulación de problemas de las ediciones anteriores. Es el resultado de la experiencia docente y de la interacción con sus principales destinatarios, los alumnos. Esta edición también se ha enriquecido incluyendo los diseños en parcelas divididas, nuevos ejercicios y la inclusión, en la sección de Ejercicios Resueltos, de soluciones basadas en el uso del paquete estadístico InfoStat para ejercicios seleccionados de los capítulos 4, 8, 9 y 10. Como en otras ediciones, hemos incorporado varias sugerencias de distintos colegas que, en distintas universidades argentinas, utilizan este material como soporte de sus cursos de grado.

Córdoba, Argentina, 2005

Índice de Contenidos

Índice de Definiciones

1

Estadística Descriptiva

Introducción

El registro de observaciones es una práctica común en el marco de la investigación. Estas observaciones surgen como resultado de un proceso de observación bajo condiciones dadas o de un proceso experimental.

Si, por ejemplo, se registraran las temperaturas mínimas diarias ocurridas en la década del 80, suponiendo un total de 3650 días, podríamos pensar que existió un proceso natural cuya realización definió la temperatura efectivamente registrada en cada uno de los 3650 días. Situaciones como ésta conducen a los conocidos **estudios observacionales**.

En otras circunstancias, las observaciones son el resultado de la provocación de un fenómeno, o experimento, bajo condiciones controladas. A modo de ejemplo, se podría considerar la aplicación de distintos insecticidas en bandejas con 100 insectos, en cada una de las cuales se registra el número de insectos muertos. Situaciones como éstas son conocidas como **estudios experimentales.**

Generalmente la información registrada en un proceso de observación es tratada, en un primer momento, con el objetivo de describir y resumir sus características más sobresalientes. Esto se conoce como *estadística descriptiva* y generalmente se basa en el uso de tablas y gráficos, y en la obtención de medidas resumen. El objetivo de este capítulo es reconocer la población y las variables relevantes en un proceso de observación o de experimentación, caracterizar y describir muestras de las poblaciones mediante medidas resumen, tablas de frecuencias y representaciones gráficas y conocer algunas metodologías de extracción de muestras.

Antes de abordar el problema de describir un conjunto de observaciones se verán algunos conceptos básicos que permiten la introducción de los procedimientos estadísticos.

Población

Definición 1.1: Población

Una **población** es un conjunto de elementos acotados en un tiempo y en un espacio determinados, con alguna característica común observable o medible.

Desde el punto de vista agronómico:

1. ¿A qué elementos hace referencia la definición? Los elementos considerados podrían ser días, animales, semillas, plantas, personas o localidades de una cierta región.
2. ¿Por qué acotar en tiempo y espacio? Dependiendo de los intereses en juego, suele ser necesario recortar el problema, o especificar claramente los alcances o fronteras del problema en estudio, ya que dentro de estos márgenes todo lo que se diga o afirme tendrá validez, y fuera de ellos no. Por ejemplo, consideremos el hecho de la estacionalidad de las precipitaciones dentro del año, y la existente entre años. Se conoce acabadamente que existen grupos de años secos y grupos de años húmedos. Más aún, que su alternancia tiene cierta frecuencia de ocurrencia. Por ello cuando estudiemos las precipitaciones acumuladas durante el mes de diciembre, será necesario especificar a qué grupo de años estamos refiriéndonos, para que lo que se analice pueda ser correctamente interpretado. El término espacio, por otro lado, puede tener en la práctica distintas connotaciones, cuestión que con el tiempo (desde el punto de vista cronológico) no ocurre. Así el espacio puede denotar una región, un volumen determinado, un lote, etc.

Definición 1.2: Tamaño poblacional

Si la población es finita, diremos que el **tamaño poblacional** es el número de elementos de la misma y lo denotaremos con **N**.

Muestra

Generalmente es imposible o impracticable examinar alguna característica en la población entera, por lo que se examina una parte de ella y en base a la información relevada en esa porción se hacen inferencias sobre toda la población.

Definición 1.3: Muestra

Se entiende por **muestra** a todo subconjunto de elementos de la población.

Definición 1.4: Unidad muestral

Una **unidad muestral** es el elemento o entidad de la muestra.

Definición 1.5: Tamaño muestral

Tamaño muestral es el número de elementos de la población que conforman la muestra y se denota con ***n***.

El problema es cómo debe ser seleccionada esa parte de la población que proveerá la información acerca de la o de las características buscadas de manera tal que puedan obtenerse conclusiones.

Vale la pena hacer una reflexión acerca del comentario, que respecto del tamaño muestral, hace uno de los más conocidos estudiosos del muestreo.

"Es clásico (y cómico) el personaje que después de pasar 10 días en un país extranjero está en condiciones de criticar la industria, reformar su sistema político, etc. Pero en realidad la diferencia que existe entre este personaje y el estudioso de ciencias políticas, que vive 20 años en ese país dedicado a estudiarlo, es que el primero basa sus conclusiones en una muestra mucho más pequeña y es menos consciente de su ignorancia" (Cochran, 1981).

En este capítulo se presentan algunas técnicas para la obtención de muestras de una población y las formas principales de resumir la información que éstas proveen. En los capítulos siguientes se verá cómo, a partir de los resúmenes muestrales, se puede estimar o inferir acerca de los parámetros distribucionales (estadística inferencial).

Variables

Las **observaciones o mediciones** sobre los elementos de una población constituyen la materia prima con la cual se trabaja en Estadística.

Para que dichas observaciones puedan ser tratadas estadísticamente deben estar expresadas o poder ser reexpresadas en términos numéricos.

Aunque sea obvio, se destaca que la característica de interés a observar o medir en cada elemento de la población debe ser la misma, en tanto que se espera que no asuma el mismo valor en cada uno de los elementos que la conforman.

Aquellas características que van cambiando en su estado o expresión entre los elementos de la población se denominan **"variables"**, mientras que aquellas que no cumplen esta condición son llamadas **"constantes".**

Definición 1.6: Variable

Una **variable** es una característica, propiedad o atributo, con respecto a la cual los elementos de una población difieren de alguna forma.

Para denotar a una cierta variable se utilizan letras mayúsculas, y con la misma letra *en minúscula* se hace referencia a un valor en particular observable en un elemento de la población, y al que se suele llamar *dato*. Así, por ejemplo, si X denota el número de semillas germinadas en un conjunto de bandejas de germinación, x denotará el número de semillas germinadas observadas en una de aquellas bandejas, siendo utilizado un subíndice para hacer referencia a un valor en particular. Así, x_{20} representa el número de semillas germinadas observadas en la bandeja número 20. Esta notación se suele generalizar, utilizando como subíndices letras minúsculas desde la *i* en adelante y luego indicando el rango de posibles valores que puede adoptar el subíndice para establecer cuántos datos se consideran en el problema.

A modo ilustrativo se presentan algunos ejemplos de notación con subíndices:

a) x_i, i=1,...,6 hace referencia taxativamente a los valores observados x_1, x_2, x_3, x_4, x_5, y x_6, no interesando otros si existieran.

b) x_i, i=1,... en este caso *i* puede valer a partir de 1 en adelante y hasta infinito.

c) x_i, i=0,1,... en este caso *i* puede valer desde cero hasta infinito.

Nota: En la práctica el término infinito, simbolizado por ∞, significará "valores inconmensurables" (negativos o positivos), sea para el subíndice (como en los casos b y c) como para los datos propiamente dichos (por ejemplo $-\infty < x_i < \infty$).

A fines ilustrativos, suponga que en la década de 1980 se registraron las temperaturas mínimas de los 3650 días. Siguiendo con la notación introducida, X hace referencia a las temperaturas mínimas en la década '80 y x_i , i=1,...,3650 a las efectivamente registradas. En particular, x_{112} denotará el valor de temperatura mínima registrado en el día 112 del período considerado; así, si en dicho día la temperatura mínima fue de -3.2 grados centígrados, escribiremos x_{112} = -3.2, y de esta forma se indica la temperatura de cualquier día en particular. De una manera general se suele denotar a un conjunto de ***n*** observaciones por $\{x_1, x_2,...,x_n\}$, donde x_n hace referencia al último término de la serie de datos. En el ejemplo anterior, *n* es 3650.

Tipos de variables

Se llamará **variable continua** a aquella característica cuyas observaciones pueden asumir cualquier valor dentro de un intervalo. En estos casos el conjunto de posibles valores es no numerable[1]. En otras palabras, existe una cantidad infinita de posibles valores para los resultados de la variable. Se puede describir el conjunto de posibles valores de una variable continua de distintas formas. Se suele seguir la siguiente convención:

a) Un intervalo es cerrado si sus extremos pertenecen al mismo, lo que se denotará con corchetes, por ejemplo, [a, b] denota al conjunto de todos los x tal que $a \leq x \leq b$.

b) Un intervalo es abierto si sus extremos no pertenecen al mismo, lo que se denotará con paréntesis, por ejemplo, (a, b) denota al conjunto de todos los x tal que $a < x < b$.

c) Un intervalo es semi-cerrado (o semi-abierto) si uno de sus extremos no pertenece al mismo, lo que se denotará con el corchete y el paréntesis que corresponda. Por ejemplo, (a, b] denota al conjunto de todos los x tal que $a<x\leq b$, en tanto $[a, b) = \{x : a \leq x < b\}$.

Es necesario no confundir el tamaño (cardinalidad) del conjunto observado, con el del conjunto de posibles valores a observar. El primero puede ser finito (es decir, es posible establecer cuántos elementos lo conforman) en tanto el segundo puede ser infinito. En el caso de las temperaturas mínimas que se presentó, el rango de posibles valores podría ser $-5°C \leq x \leq -1°C$, un intervalo continuo, y por lo tanto con infinitos valores posibles de ser observados, en tanto que el tamaño de la población considerada es de 3650 días.

En la práctica ocurre que los instrumentos de medición producen un redondeo del verdadero valor que presenta el elemento a medir, según la precisión que acrediten. Pero no por ello se deberá decir que el rango de posibles valores es finito. Si el termómetro con que se realizan las mediciones de las temperaturas mínimas mide en °C con 2 decimales de precisión, entre $1°C \leq x \leq 5°C$ existirían (500-100) + 1 = 401 datos posibles, no obstante esta variable es de tipo continua.

Se llaman **variables discretas**, en contraposición a las variables continuas, a aquellas características que asumen un número finito o infinito numerable de valores posibles. Así, las variables discretas surgen de *conteos*, como por ejemplo el número de días hasta la germinación del 50% de las semillas de una bandeja, número de colonias de microorganismos sobre plantas enfermas, el número de frutos de un árbol, el número de mazorcas en plantas de maíz, etc.

Se llaman **variables categóricas**, en contraposición a las variables cuantitativas, a

[1] Se dice que un conjunto es infinito numerable si cada uno de sus elementos se asocia biunívocamente con un número natural, en caso contrario se dice que el conjunto es no numerable.

aquellas cuya escala de medida es un conjunto de categorías. Entre ellas podemos distinguir al menos:

a) *Categóricas nominales,* como la orientación de los vientos, que se podrían considerar como “Norte”, “Sur”, “Este”, “Oeste”; el color del tegumento de las semillas, el sexo, etc.

b) *Categóricas ordinales,* como el grado de ataque de una virosis vegetal que puede ser "severo", "moderado" o "leve".

Es importante señalar que las variables continuas se pueden “discretizar” (por ejemplo tomando intervalos) y así ser tratadas como discretas o, cuando una variable discreta asume una gran variedad de valores, como podría ser el caso de contar el número de pulgones en hojas de trigo, ésta puede ser tratada como una variable continua.

Muestreo aleatorio simple

Definición 1.7: Muestreo aleatorio simple

El **muestreo aleatorio simple** es el método de selección de ***n*** unidades de una población de tamaño **N** de tal modo que cada una de las muestras posibles tenga la misma oportunidad de ser elegida (Cochran,1981).

Para obtener una muestra aleatoria simple se enumeran las unidades de la población de 1 a **N** y posteriormente se extrae una serie de *n* números aleatorios entre 1 y **N** (tarea que se puede realizar usando una tabla de números aleatorios o mediante un programa de computación que produce una tabla semejante). Las unidades cuya numeración coincide con la serie de números seleccionados conformarán la muestra aleatoria. En este esquema muestral si una unidad muestral fue previamente seleccionada, entonces no puede ser seleccionada nuevamente. En cada extracción el proceso debe garantizar *la misma oportunidad de selección a todos y cada uno de los elementos* que no hayan sido seleccionados aún. Por este método existen $C_n^N n!$ formas posibles de obtener *n* elementos de entre N. No obstante, sólo existen C_n^N muestras (conjuntos diferentes) todas con igual oportunidad de ser extraídas. La probabilidad de cada muestra es entonces igual a $\frac{1}{C_n^N}$. El método recibe también el nombre de muestreo **sin restitución** porque en la muestra no puede aparecer el mismo elemento repetido, es decir, que una vez que un elemento ha sido extraído no es restituido y por lo tanto no está disponible para la elección del próximo elemento de la muestra.

Por ejemplo, se tiene una población de seis elementos identificados como: **a**, **b, c, d, e, f** y se desea saber cuántas muestras posibles de tamaño 3 se pueden tomar de la misma utilizando un esquema de muestreo sin restitución. Si el tamaño poblacional es N = 6 y el de la muestra es *n* = 3, entonces el número de muestras posibles sin restitución es:

$$C_3^6 = \frac{6!}{(6-3)!\ 3!} = \frac{720}{36} = \frac{120}{6} = 20$$

La muestras posibles son las siguientes:

a, b, c	a, b, d	a, b, e
a ,b, f	a, c, d	a, c, e
a ,c, f	a, d, e	a, d, f
a, e, f	b, c, d	b, c, e
b, c, f	b, d, e	b, d, f
b, e, f	c, d, e	c, d, f
c, e, f	d, e, f	

En los puntos que siguen, cuando se haga referencia a muestra, se considerará solamente a la obtenida a partir de un muestreo aleatorio simple **con restitución.** En este tipo de muestreo la cantidad de formas posibles de extraer *n* elementos desde una población de tamaño N es igual a N^n. Por ejemplo, si una población tiene 2 elementos identificados con **a** y **b** y se quiere saber cuantas formas se tiene de extraer tres elementos, estas son 2^3=8 y están dadas por: {aaa, aab, aba, baa, bba, bab, abb, bbb}. Nótese que aab, aba, y baa contienen los mismos elementos, por lo cual éstas constituyen la misma muestra (dos conjuntos con iguales elementos son indistinguibles) luego el total de muestras posibles es menor que N^n pero en este caso las muestras no son todas igualmente probables (ver Capitulo 2).

Resumen de la información muestral

Al registrar los resultados de un estudio observacional o experimental, se obtiene un número de observaciones que puede ser muy grande y su simple listado es de poca relevancia en el sentido interpretativo. Aunque a partir de dichos registros se puede encontrar la respuesta buscada, no están ordenados de manera tal que adquieran significado para el investigador. Es por esto deseable presentar las observaciones en forma resumida.

A los fines de ordenar, resumir y presentar la información, se utilizan tablas y gráficos apropiados para cada tipo de variable (variables numéricas, continuas o discretas, o bien, variables no numéricas o de naturaleza categórica), por lo que trataremos las distintas situaciones por separado.

Tablas de distribución de frecuencias y gráficos para variables discretas

Una tabla de distribución de frecuencias posee una columna que contiene los diferentes valores que toma la variable en estudio y otra columna que indica la **frecuencia absoluta**.

Definición 1.8: Frecuencia absoluta

Se denomina **frecuencia absoluta** al número de veces que el valor de la variable se repite en el conjunto de datos.

Simbolizando con x_i, $i = 1,2,3....m$, a los distintos valores observados para la variable X y por n_i a la frecuencia absoluta del valor x_i, podemos agrupar los datos en una tabla de frecuencias de la siguiente manera :

x_i	n_i
x_1	n_1
x_2	n_2
.	.
.	.
.	.
x_m	n_m
	n

con $n = n_1 + n_2 + n_3 + ... + n_m = \sum_{i=1}^{m} n_i$, representando el número total de observaciones.

Generalmente en una tabla de distribución de frecuencias no sólo se muestran las frecuencias absolutas, sino que también se incluyen las frecuencias relativas y las frecuencias acumuladas (absolutas y relativas), como es descripto en el ejemplo a continuación:

Ejemplo 1.1

Un experimento consistió en contar el *número de flores* por planta de una muestra

n = 50 plantas. Los valores resultantes del conteo fueron los siguientes:

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

Los datos así presentados son de difícil interpretación, por lo que conviene resumirlos como en la siguiente tabla:

Tabla 1.1: Tabla de distribución de frecuencias para la variable número de flores por planta.

i	x_i	n_i	N_i	f_i	F_i
1	0	1	1	0.02	0.02
2	1	1	2	0.02	0.04
3	2	3	5	0.06	0.10
4	3	6	11	0.12	0.22
5	4	4	15	0.08	0.30
6	5	4	19	0.08	0.38
7	6	9	28	0.18	0.56
8	7	8	36	0.16	0.72
9	8	7	43	0.14	0.86
10	9	4	47	0.08	0.94
11	10	3	50	0.06	1.00

En esta tabla se puede ver que el número total de datos es 50, que las plantas con menos de 3 flores y con más de 9 son poco frecuentes y que plantas que tienen entre 6 y 8 flores son las más frecuentes. Esta tabla de frecuencias se construye de la siguiente forma:

a) En la primera columna se colocan los valores de $i = 1,...,m$, donde m es el número de diferentes valores que asume la variable X.

b) En la segunda columna se colocan los valores observados x_i, diferentes entre sí, de la variable X que representa el número de flores por planta.

c) En la tercera columna el número de veces que aparece cada valor, o sea, la frecuencia absoluta (n_i).

d) En la cuarta columna se observan las frecuencias absolutas acumuladas, denotadas por N_i. Éstas se definen como el valor que surge de la acumulación por fila de las correspondientes frecuencias absolutas, o sea:

$N_1 = n_1$; $N_2 = n_1 + n_2$; $N_3 = n_1 + n_2 + n_3$; y así sucesivamente.

En general, se expresan las frecuencias absolutas acumuladas de la siguiente forma:

$$N_k = n_1 + n_2 + \ldots\ldots + n_k = \sum_{i=1}^{k} n_i \text{ , con } k \leq m$$

e) En la quinta columna se calculan las frecuencias relativas, denotadas por f_i, esto es: $f_i = \frac{n_i}{n}$, $\forall\ i = 1,\ldots,m$.

O sea, la frecuencia relativa es el cociente entre cada frecuencia absoluta y el total de observaciones; en el ejemplo, $f_5 = \frac{4}{50} = 0.08$. Cada una de estas frecuencias multiplicadas por 100, indica el porcentaje con que cada valor de la variable está representado en la muestra. Así se consigue una mejor apreciación del peso que tiene cada valor de la variable en su *distribución de frecuencias*. Por ejemplo $f_5 = 0.08$, entonces $f_5 \cdot 100 = 8\%$; es decir que el 8% de las plantas en la muestra tienen 4 flores.

La frecuencia relativa asociada con un valor dado, puede ser considerada como una estimación de la probabilidad de ocurrencia de dicho valor. Una propiedad que cumplen las frecuencias relativas es que su suma es igual a 1.

f) En la sexta columna se calculan las frecuencias relativas acumuladas, sumando las frecuencias relativas de la misma manera que se sumaron las frecuencias absolutas para obtener las absolutas acumuladas. Las frecuencias relativas acumuladas serán denotadas con F_i y calculadas como:

$$F_i = \frac{N_i}{n}, \forall\ i = 1,\ldots,m$$

donde F_i es la frecuencia relativa acumulada para el i-ésimo valor de X.

¿Qué información se obtiene de la tabla de frecuencias así construida?

Los valores 6, 7 y 8 de la variable número de flores por planta, fueron los que se observaron con mayor frecuencia, 9 plantas (18%) presentaron 6 flores, 8 plantas (16%) tuvieron 7 flores, 7 plantas tuvieron 8 flores; pocas fueron las plantas sin flores (2%); el 10% de las plantas tuvieron 2 o menos flores; el número máximo de flores por planta en esta experiencia fue de 10 y sólo en el 6% de la muestra se registró este valor máximo.

Estas afirmaciones, como algunas otras, pueden obtenerse de la lectura de una tabla de frecuencias, y no son fáciles de formular a partir de los datos sin procesar, sobre todo cuando n es grande.

La información de una tabla de frecuencias también puede ser presentada gráficamente. Si en el eje de las ordenas se disponen las frecuencias absolutas o relativas y en el eje de las abscisas los distintos valores que toma la variable, se obtiene un gráfico de barras de frecuencias absolutas o relativas, respectivamente, como se muestra en la Figura 1.1.

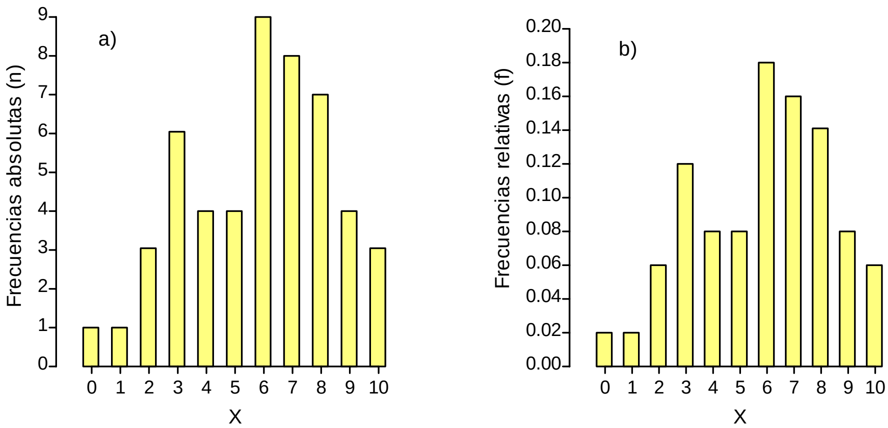


Figura 1.1: Gráfico de barras de: a) frecuencias absolutas b) frecuencias relativas.

Cuando en el eje de ordenadas se representan las frecuencias acumuladas, ya sean absolutas o relativas (N_i o F_i), el gráfico que se obtiene se muestra en la Figura 1.2.

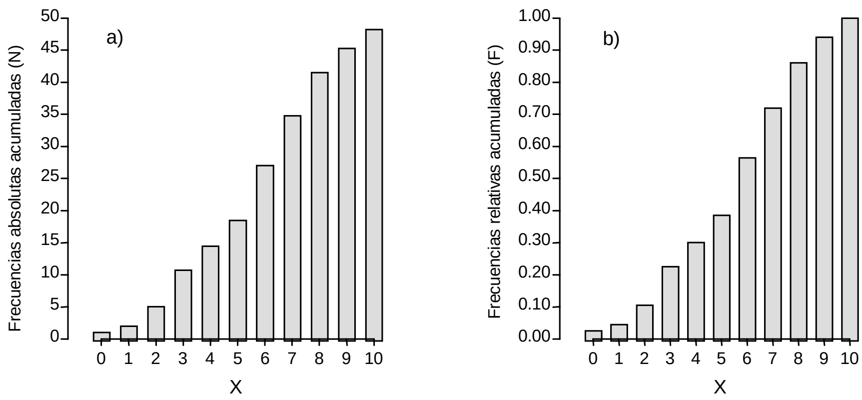


Figura 1.2: Gráfico de: a) frecuencias absolutas acumuladas b) frecuencias relativas acumuladas.

La información que presentan las figuras anteriores es equivalente a la presentada en la tabla de frecuencias y la utilización de una u otra es un problema de elección que se debe resolver según un criterio de oportunidad. No obstante, si el objetivo es visualizar la forma en la que se distribuyen los datos para seleccionar un modelo probabilístico para la variable, la representación gráfica es, seguramente, la mejor opción.

Cabe mencionar que el tratamiento de las **variables categóricas** es similar al de las variables numéricas discretas. Una tabla de frecuencias para una variable categórica se construye de la misma forma que en el caso de las variables discretas.

Para representar gráficamente datos categóricos, en general, se utilizan los gráficos de barras o los diagramas en torta. Ejemplos de este tipo de representación pueden consultarse en el apartado correspondiente a Representaciones Gráficas, al final de este capítulo.

Tablas de distribución de frecuencias y gráficos para variables continuas

Para describir la distribución de frecuencia correspondiente a una variable continua, es indispensable agrupar los valores registrados mediante un conjunto de intervalos. La determinación de la cantidad y amplitud de los intervalos es arbitraria. La distribución de frecuencia debería tener entre 5 y 15 intervalos puesto que, si no hay suficientes intervalos habrá demasiada concentración de datos, y si hay demasiados puede suceder que algunos no contengan observaciones.

Una forma no arbitraria de obtener el número de intervalos es calcularlo como $\log_2(n+1)$; ($\log_2$ se puede obtener como $\log_2(x)=\log_a(x)/\log_a(2)$).

Una regla práctica para definir la amplitud de los intervalos consiste en:

a) Calcular el recorrido o rango de la variable tomando la diferencia entre el mayor y el menor valor.

b) Dividir el recorrido por la cantidad de intervalos que se quiere tomar. El resultado de ese cociente es la amplitud que tendrá cada intervalo.

La frecuencia absoluta correspondiente a cada intervalo es la cantidad de valores observados de la variable en dicho intervalo.

Ejemplo 1.2

Se toma una muestra de 100 espigas de trigo y en cada una de ellas se registra la longitud en cm. En este caso la variable X es longitud de espiga. Los resultados son los que se presentan a continuación:

10.3	12.8	8.3	6.9	10.2	11.1	11.9	8.7	9.5	6.9
7.3	8.1	6.4	16.0	12.9	8.6	10.6	9.3	14.1	12.8
11.6	8.7	7.9	8.6	8.9	6.8	7.9	11.6	10.3	11.8
9.7	12.8	13.1	6.3	8.4	8.9	10.6	11.3	7.8	14.6
8.7	8.5	9.3	10.6	11.4	13.7	8.5	9.7	10.1	10.8
6.9	7.5	15.0	9.3	10.0	10.6	11.3	11.4	9.9	7.8
10.6	11.8	10.5	10.7	10.6	14.9	13.2	10.9	10.6	11.1
9.9	6.7	7.8	10.9	10.6	11.3	9.8	9.6	15.0	11.0
12.3	12.6	9.5	6.3	8.7	10.5	14.0	13.6	10.1	6.9
8.6	7.0	6.8	11.4	13.2	6.9	7.9	10.3	10.9	11.3

Aplicando la regla para calcular el número de intervalos se tiene que $\log_2(100+1) \cong 7$. Para el cálculo de la amplitud de los intervalos en este ejemplo se tiene:

a) El recorrido es 16 - 6.3 = 9.7

b) Como el número de intervalos a tomar es 7, la amplitud resultante es 9.7/7 = 1.39.

Para facilitar la construcción de la tabla se redondea de 1.39 a 1.4.

Por lo tanto, los conjuntos de valores de 6.3 a 7.7 cm, 7.7 a 9.1,..., 14.7 a 16.1 cm, constituyen los 7 intervalos para agrupar los valores observados de la variable longitud de espiga.

El límite inferior del intervalo se simbolizará como LI y el límite superior como LS. El valor promedio entre los límites del intervalo se llama "punto medio" del intervalo o "marca de clase" y se denota por m_i.

Para construir la tabla de frecuencias de una variable continua se procede de igual forma que para una variable discreta, es decir, se tabulan las frecuencias absolutas, las frecuencias absolutas acumuladas, las frecuencias relativas y las frecuencias relativas acumuladas en cada intervalo. Para los datos del Ejemplo 1.2 se obtiene la Tabla 1.2.

Tabla 1.2:Tabla de distribución de frecuencias correspondiente al Ejemplo 1.2.

LI, LS	n_i	m_i	N_i	f_i	F_i
[6.3,7.7]	14	7.0	14	0.14	0.14
(7.7, 9.1]	20	8.4	34	0.20	0.34
(9.1, 10.5]	18	9.8	52	0.18	0.52
(10.5, 11.9]	29	11.2	81	0.29	0.81
(11.9, 13.3]	10	12.6	91	0.10	0.91
(13.3, 14.7]	5	14.0	96	0.05	0.96
(14.7, 16.1]	4	15.4	100	0.04	1.00
	100				

¿Qué significa n_i, $\forall\ i$, en esta tabla? Por ejemplo, para el intervalo (9.1,10.5] n_i es igual a 18, pues 18 de los valores que se observaron en la muestra están en dicho intervalo. Se dejan para el lector otras interpretaciones de esta tabla.

Cuando se grafican en el eje de las ordenadas las frecuencias absolutas o relativas y en el eje de las abscisas los intervalos en los que se encuentran valores de la variable, el gráfico obtenido se denomina **histograma de frecuencias absolutas o relativas**, respectivamente.

Como dentro de cada intervalo existen varios valores de la variable, se construirá una barra de altura igual a la frecuencia de dicho intervalo y ancho igual a la amplitud del mismo.

Otro gráfico que puede utilizarse es el **polígono de frecuencias.** Este es un gráfico construido al unir los puntos medios de los extremos superiores de las barras de un histograma por segmentos de recta, como puede observarse en la Figura 1.3.

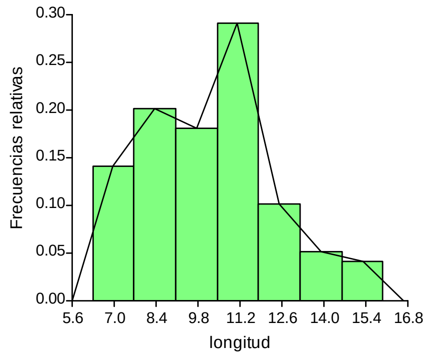


Figura 1.3: Histograma y polígono de frecuencias relativas.

Para las frecuencias absolutas acumuladas (o relativas acumuladas), los gráficos que se usan son **histogramas y/o polígonos de frecuencias absolutas acumuladas** (o relativas acumuladas). Estos gráficos se confeccionan poniendo en el eje de las ordenadas las frecuencias absolutas (o relativas) acumuladas y en el eje de las abscisas los intervalos. Para construir el polígono se unen los puntos de la derecha del extremo superior de cada barra, como se muestra en la siguiente figura.

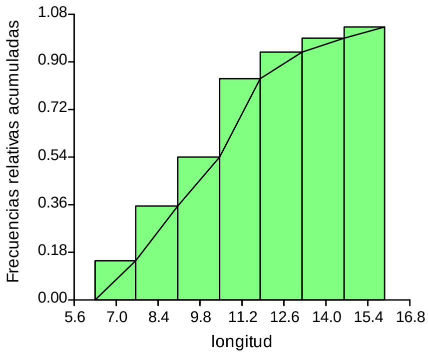


Figura 1.4: Histograma y polígono de frecuencias acumuladas relativas.

Más adelante se verá que el polígono de frecuencias permite aproximar la función de densidad de una variable aleatoria, mientras que el polígono de frecuencias relativas acumuladas da una aproximación de su función de distribución.

Medidas resumen de la información muestral

Los aspectos relevantes de una muestra son generalmente descriptos usando medidas de posición y de dispersión. Estas medidas, como se verá luego, sirven para aproximar los parámetros de posición y de dispersión de una variable.

Medidas de posición

Las medidas de posición son funciones de los datos de una muestra que miden, según diferentes criterios, el centro de la distribución de frecuencias en la muestra; también son llamadas medidas de tendencia central. Las más usadas para este fin son: la media muestral ($\bar{x}$), los cuantiles (x_p), la mediana muestral (me) y el modo muestral (mo).

Definición 1.9: Media muestral o promedio

Si x_1, x_2,....., x_n constituyen una muestra aleatoria de tamaño *n*, luego la **media muestral o promedio en la muestra** se define como:

$$\overline{x} = \sum_{i=1}^{n} \frac{x_i}{n}.$$

Definición 1.10: Cuantil muestral

Si x_1, x_2,....., x_n constituyen una muestra aleatoria de tamaño *n* entonces el **cuantil p** de su distribución de frecuencias muestral es el valor que en la muestra ordenada en forma ascendente ocupa la posición [p x n] con p tal que 0<p<1.

Nota: en la definición anterior [.] indica tomar la parte entera de la expresión incluida.

En otras palabras, el cuantil *p* es aquel valor observado de la variable, en la muestra, tal que el número de valores menores o iguales a él constituyen la proporción p del número total de observaciones en la muestra.

Esta definición está basada en una forma de obtener los cuantiles y permite calcularlos adecuadamente para tamaños muestrales medianos o grandes.

Nota: Es frecuente referirse a percentiles como sinónimo de cuantiles. Su diferencia está en que en el primero se especifica el porcentaje y en el segundo la proporción.

Definición 1.11: Mediana muestral

Si x_1, x_2,....., x_n constituyen una muestra aleatoria de tamaño *n* entonces la **mediana** muestral es el cuantil 0.50 de su distribución de frecuencias muestral.

La mediana, al igual que otros cuantiles, puede ser obtenida a partir del polígono de distribución de frecuencias relativas acumuladas o, como regla práctica, ordenando los valores de la variable de menor a mayor y tomando el valor central en el caso en que *n* es impar o el promedio de los dos valores centrales cuando *n* es par.

La mediana es una medida de tendencia central que no es afectada por valores extremos o atípicos como lo es la media. A esta propiedad se la conoce como robustez.

Definición 1.12: Moda muestral

Si x_1, x_2,....., x_n conforman una muestra aleatoria, la **moda muestral** es el valor de la variable que ocurre con mayor frecuencia.

Para el caso de variables continuas la moda es la marca de clase que posee mayor frecuencia.

Pueden existir uno o más valores modales. Para distribuciones simétricas unimodales, la media, la mediana y la moda corresponden al mismo valor.

Para los datos presentados en el Ejemplo 1.2 se tiene que: $\bar{x}$ =10.176; mediana=10.3; moda=11.2.

Medidas de dispersión

Para describir la dispersión de los elementos de una muestra, las medidas que se encuentran con mayor frecuencia en la literatura técnica (Steel y Torrie, 1985; Berenson *et al.*, 1983) son el **rango muestral**, el **rango intercuartílico**, la **varianza,** el **desvío estándar o desviación estándar muestral** y el **coeficiente de variación muestral.** Se verá en capítulos posteriores cómo se utilizan estas medidas para inferir acerca de los parámetros correspondientes de la distribución.

Definición 1.13: Rango muestral

Dada una muestra aleatoria x_1, x_2,....., x_n, el **rango muestral** se define como $r = x^{(n)} - x^{(1)}$, donde $x^{(n)}$ y $x^{(1)}$ corresponden a los valores máximo y mínimo en la muestra respectivamente.

Definición 1.14: Varianza muestral

Si $x_1, x_2, \ldots, x_n$ conforman una muestra aleatoria, la **varianza muestral** es una función de los desvíos de cada x_i respecto a la media muestral $\bar{x}$ que tiene la siguiente expresión:

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2.$$

Definición 1.15: Desviación Estándar muestral

Dada una muestra aleatoria $x_1, x_2, \ldots, x_n$ con varianza muestral S^2, la **desviación estándar muestral** se define como $S = \sqrt{S^2}$.

Tanto la varianza como la desviación estándar muestrales miden la dispersión de los valores observados con respecto a la media de la muestra. La diferencia entre ellas es que mientras la primera está dada en unidades al cuadrado (si la variable se mide en cm la varianza se mide en cm^2), la segunda tiene la misma unidad de medida que la media. En este sentido para la presentación de resultados suele preferirse a esta última medida.

Cuando se desea hacer referencia a la relación entre el tamaño de la media y la variabilidad de las observaciones, se usa el coeficiente de variación muestral.

Definición 1.16: Coeficiente de variación muestral

Dada una muestra aleatoria $x_1, x_2, \ldots, x_n$ con media $\bar{x}$ y desviación estándar S, el **coeficiente de variación muestral** se define como:

$$CV = \frac{S}{\bar{x}} \cdot 100.$$

Esta medida es adimensional y permite en consecuencia comparar la variabilidad de características medidas en diferentes escalas. Por ejemplo, si se comparan dos índices para medir nivel de ataque de pulgones y ambos están basados en técnicas completamente diferentes, que dan puntajes cuyas unidades de medida son distintas, se dirá que el índice que tenga menor coeficiente de variación es el menos variable.

Para los datos presentados en el Ejemplo 1.2 se tiene que: S^2=5.063; S=2.25; CV=22.11%.

Otros tipos de muestreos

Existen situaciones en las que el muestreo aleatorio simple suele no ser óptimo. A continuación se presentan, brevemente, algunas técnicas alternativas para la obtención de muestras. En la práctica profesional del muestreo en ocasiones es necesaria la combinación de dos o más de estas técnicas para obtener la mayor información al menor costo.

Muestreo Estratificado

Se usa cuando la característica en estudio no presenta una distribución aleatoria sobre las unidades muestrales. Los estratos (subconjuntos de unidades muestrales) deben tomarse de manera tal que los valores de la variable sean más homogéneos dentro de los mismos que entre ellos. Al existir un patrón de estratificación de las unidades respecto a la variable en estudio, se obtiene una caracterización más precisa de dicha variable, con muestreo estratificado que aplicando un m.a.s.

El muestreo estratificado consiste en extraer una muestra aleatoria dentro de cada estrato y luego combinar la información proveniente de los distintos estratos, ponderada por el tamaño de los mismos. Por ejemplo, para obtener la media muestral de un muestreo estratificado es necesario combinar las medias por estrato, ponderándolas por el tamaño de los mismos. La medida resultante se denomina **promedio ponderado.** Esta medida no es de uso exclusivo en el contexto del muestreo por estratos sino de uso general y su finalidad es combinar otras medidas teniendo en cuenta ponderadores.

Definición 1.17: Promedio ponderado

Si w_i es el peso o ponderación asociada con el valor mustral x_i y x_1, x_2,....x_n es una muestra aleatoria de tamaño n, el **promedio ponderado muestral** $\bar{x}$ ponderada es:

$$\bar{x}_{ponderada} = \frac{\sum_{i=1}^{n} x_i w_i}{\sum_{i=1}^{n} w_i}.$$

Aplicando la definición anterior al muestreo estratificado, la media obtenida por esta técnica está dada por:

$$\bar{x}_{ponderada} = \frac{\sum_{i=1}^{k} \bar{x}_i n_i}{\sum_{i=1}^{k} n_i}$$

Donde $\bar{X}_i$ es la media del i-ésimo estrato y n_i es el tamaño del i-ésimo estrato, para todo *i*. La varianza, en este esquema de muestreo, se obtiene de manera semejante a la media ponderada, sustituyendo en la expresión dada anteriormente las $\bar{x}_i$ por S_i^2 y los n_i por $(n_i\text{-}1)$.

Muestreo por Conglomerados

Se aconseja su uso cuando por razones de costos, de practicidad u otras consideraciones no es conveniente enumerar y acceder a cada unidad de muestreo, y éstas se pueden agrupar en subconjuntos, denominados conglomerados, de forma tal que haya heterogeneidad entre las unidades de un mismo conglomerado y homogeneidad entre conglomerados. El concepto de homogeneidad entre conglomerados se refiere a que las medidas que se pueden calcular para cada conglomerado difieren poco de conglomerado en conglomerado.

Al existir un patrón de conglomerados de las unidades muestrales, se obtiene una estimación más precisa si se muestrean aleatoriamente un número determinado de conglomerados y se censan todas las unidades muestrales que los constituyen que si se realizara un m.a.s.

Al igual que en el muestreo por estratos, la información de los distintos conglomerados incluidos en la muestra debe combinarse para obtener la información deseada; una forma es promediando la información de cada uno de ellos. Por ejemplo, si se desea estimar la producción de un cultivo en un departamento de la Provincia de Córdoba, es necesario visitar una gran cantidad de campos. En este caso, hacer un m.a.s. es muy costoso ya que se deben recorrer muchos kilómetros para recaudar la información. Dividiendo el departamento en áreas o conglomerados y seleccionando aleatoriamente algunos de ellos, y luego censándolos (es decir, se visita a todos los productores), se obtendrá la misma información con menor costo operativo.

Esta práctica permite revisar más unidades muestrales con el mismo esfuerzo o costo que con un m.a.s., ya que al pertenecer todos los campos a un área determinada, las distancias a recorrer son cortas.

Otro ejemplo de muestreo por conglomerados, algo distante de las ciencias agropecuarias, es el que aparece en el análisis textual cuando se quiere conocer la frecuencia de distintos vocablos en la obra de un escritor. Carecería de sentido tratar de identificar y numerar todas las palabras de una obra literaria para luego hacer un muestreo aleatorio simple. Sería mucho más práctico, tomar páginas al azar y revisar todas las palabras de esa página.

Debe quedar claro que este método presupone que la información dentro de cada conglomerado es más variable que entre conglomerados, de lo contrario no es aconsejado.

Muestreo por Captura y Recaptura

Esta es una técnica especializada de muestreo que se aconseja cuando el objetivo es conocer el tamaño de una población cuyos elementos, por alguna razón, no están fijos en el espacio. Un ejemplo típico es la estimación de tamaño de poblaciones de animales silvestres.

Esta técnica se basa en la igualdad de proporciones: A/B = C/D, donde D es el tamaño de la población a determinar. Se extrae una muestra de tamaño C, se marcan de alguna forma los elementos que la conforman y luego se devuelven o reintegran a la población. En una segunda instancia y suponiendo que los elementos de la muestra de tamaño C se han mezclado uniformemente en la población, se toma una nueva muestra de tamaño B y se registra como A al número de elementos que se encuentran marcados. Luego D = B*C/A.

Este procedimiento se repite un número suficiente de veces y se promedian los resultados a fin de obtener una mejor estimación del tamaño de la población: $\overline{D}$.

Supongamos por ejemplo que se desea saber la cantidad de peces de una especie en una laguna. Para ello, se extrae una muestra de peces de tamaño C = 300, se los marca, y luego se los reintegra a la laguna. Posteriormente, se toma una segunda muestra de tamaño B=200 y se cuenta la cantidad de peces marcados, A = 10. Aplicando la igualdad de proporciones, se tiene que D=200 (300/10) = 6000.

Representaciones gráficas

Aunque las medidas descriptivas como la media, varianza, moda, mediana, rango, cuantiles, etc. sirven para describir de manera cuantitativa distintos aspectos de la muestra, los recursos gráficos adquieren especial importancia a la hora de presentar resultados con fines de divulgación o para explorar dicho conjunto de datos. Más allá de cualquier otro objetivo, se debe tener presente que desde la perspectiva estadística las representaciones gráficas auxilian en la visualización de la información.

Aunque el hombre es un ser predominantemente visual, acostumbrado a tomar decisiones en función de la información que recibe por esa vía, no está especialmente preparado para manejar representaciones gráficas a través de las cuales se intenta proveer información cuantitativa precisa.

Muchos estudios se han realizado para comprender el proceso de percepción visual y sobre las mejores estrategias de comunicación asociadas (Cleveland, 1984, 1985; Bahrd, 1970; Shepard, 1978; Kosslyn 1980). Las recomendaciones indican que los gráficos deben ser simples, con pocos elementos iconográficos, con escalas perfectamente identificadas y preferentemente bidimensionales. Estas recomendaciones contrastan con lo que habitualmente se ve en revistas de divulgación y en la propaganda televisiva, pero se debe tener presente el carácter no cuantitativo de la información que estas presentaciones tratan de transmitir.
Usando criterios psicométricos, Cleveland (1985) propuso una escala (Figura 1.5) para los elementos gráficos según su calidad para transmitir con exactitud las diferencias de las magnitudes que representa. La moraleja es cuanto más simple mejor.

En esta sección se presentan distintas formas de representación gráfica para las cuales se discuten sus propósitos, casos típicos de aplicación y eventualmente algunas variaciones del esquema básico. De ninguna manera se pretende hacer un tratamiento exhaustivo de los métodos gráficos ya que está fuera de los objetivos de este material pero sí se pretende que haya un conocimiento y manejo básico de los recursos gráficos para la representación de datos en Estadística.

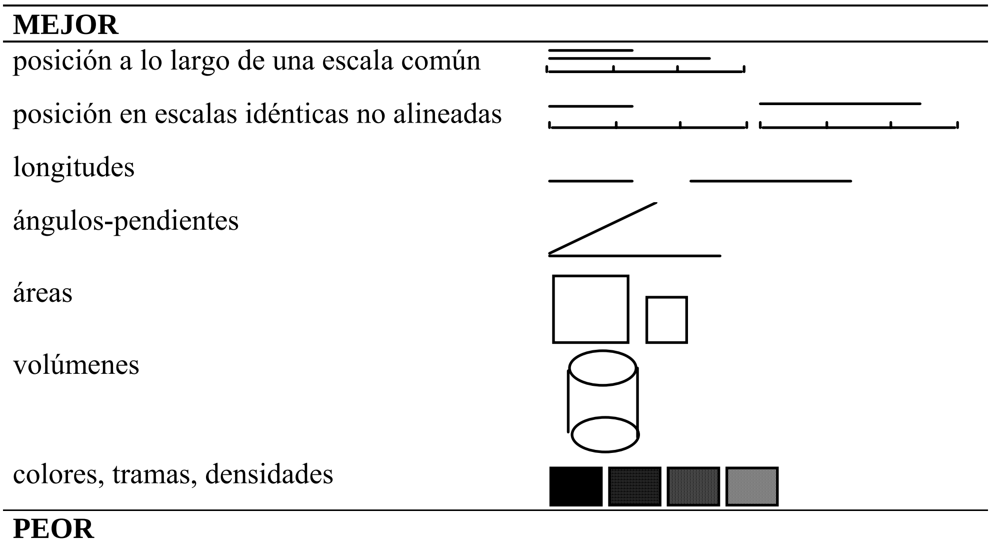


Figura 1.5: Escala de calidad de los elementos gráficos para representar magnitudes comparativamente.

Gráfico de Barras

Una aplicación muy usada de estos gráficos es la representación de frecuencias absolutas o relativas de distintas categorías en las que se pueden clasificar las unidades de observación. El ejemplo que sigue muestra la frecuencia absoluta de pulgones en distintos estadios de desarrollo. La Figura 1.6 informa que el estadio ninfa 1 es el más frecuente con 300 casos, le sigue el estadio ninfa 2 con cerca de 200 casos, el estadio ninfa 3 con menos de 100 y los adultos con menos de 30. La Figura 1.7 es una variante del gráfico anterior donde la lectura de la frecuencia es inequívoca y se podría haber suprimido por completo la escala vertical que indica la frecuencia a cambio de indicar en la leyenda el modo de leer la figura como se hace en la Figura 1.8.

Si en vez de representar frecuencias absolutas se representan frecuencias relativas, las figuras no cambian excepto que la escala en el eje de ordenadas tendrá un máximo menor o igual a 1.

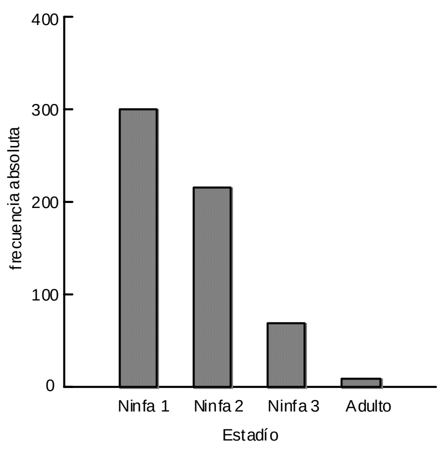


Figura 1.6: Distribución de frecuencias absolutas de pulgones según su estadio de desarrollo.

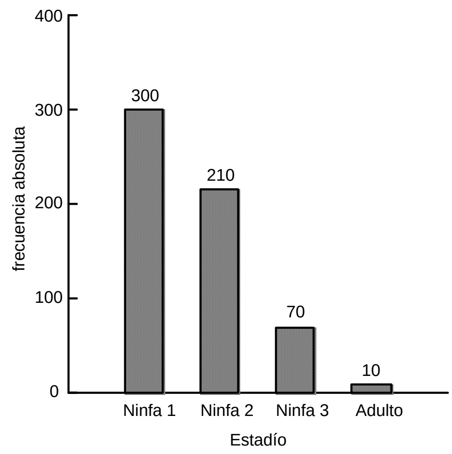


Figura 1.7: Distribución de frecuencias absolutas de pulgones según su estadio de desarrollo.

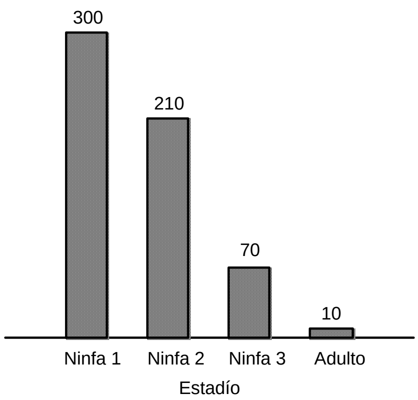


Figura 1.8: Distribución de frecuencias absolutas de pulgones según su estadio de desarrollo.

Una aplicación frecuente del gráfico de barras es la representación de los valores medios de una variable. Cada barra tendrá una altura, en la escala del eje de ordenadas, igual a la media que se quiere representar. Tomando el ejemplo anterior se puede mostrar para cada estadio de desarrollo, el peso promedio de los individuos. Además, en estos casos, es una buena práctica agregar a estas representaciones una medida de la variabilidad muestral de la media, mediante un segmento de recta colocado en la parte superior de cada barra y cuya longitud es igual al error estándar de la media, $EE = \sqrt{\sigma^2 / n}$, y será definida en el Capítulo 4. Una aplicación como esta daría como resultado un gráfico como el que muestra la Figura 1.9.

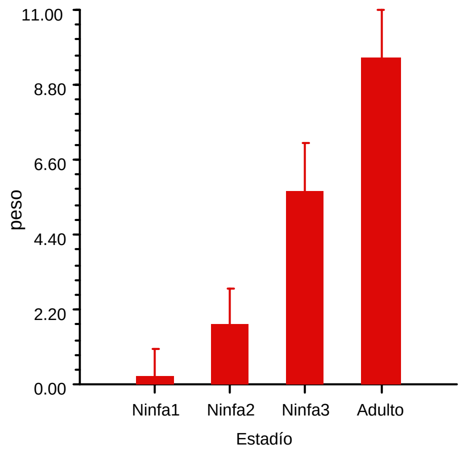


Figura 1.9: Peso promedio y error estándar de los distintos estadios de desarrollo de pulgones.

Diagramas de Torta

Una alternativa para la representación de frecuencias relativas de un conjunto de categorías es la utilización de diagramas en torta. En este caso la porción de torta que le corresponde a cada categoría representa la frecuencia relativa. La Figura 1.10 muestra la representación en diagrama de torta del ejemplo presentado en la Figura 1.7 utilizando las frecuencias relativas. Una limitante para este tipo de representaciones es el número de categorías, ya que cuando éstas son muchas, la lectura del gráfico no es simple. Hay varias variantes al esquema básico, las más frecuentes incluyen profundidad (3D o 3 dimensiones), perspectiva y la posibilidad de distinguir alguna categoría mediante la reubicación de su porción de torta en una posición más alejada del resto. Asimismo, se pueden agregar los porcentajes que cada porción de la torta representa para facilitar así su lectura.

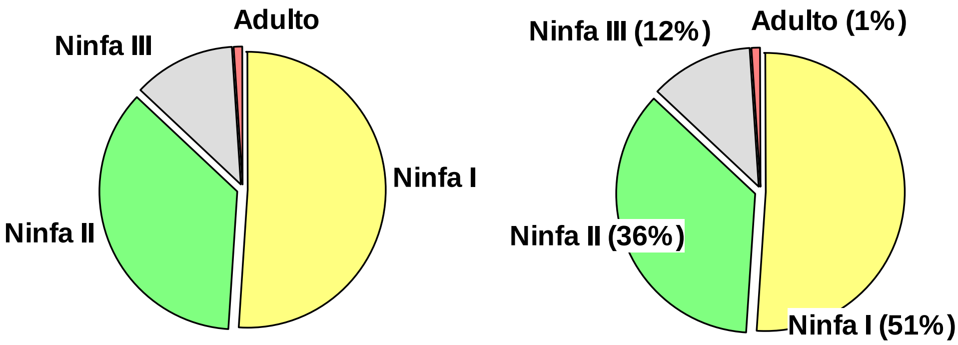


Figura 1.10: Distribución de individuos según el estadio de desarrollo sin identificación de las magnitudes que representan cada una de las porciones de la torta (a) y con la aclaración correspondiente (b).

Diagramas de Caja (´Box Plot´)

Estos gráficos tienen por objeto presentar sintéticamente los aspectos más importantes de una distribución de frecuencias.

Ejemplo 1.3

Se toman muestras aleatorias de tamaño n = 100 de cada uno de tres estadios larvales de una especie de polilla forestal. Cada individuo es pesado y los resultados se presentan en la Tabla 1.3.

Tabla 1.3: Peso (mg) de 100 larvas de cada estadio de una polilla forestal.

Estadio 1			Estadio 2			Estadio 3		
0.47	2.87	0.06	2.40	4.85	3.09	22.74	7.96	10.03
0.05	0.24	0.63	3.48	4.46	9.22	3.63	11.19	4.54
0.25	0.00	0.86	3.69	10.67	5.28	8.17	15.34	10.88
1.43	0.00	0.00	5.35	1.75	2.25	9.82	5.14	4.68
0.49	0.28	0.04	3.01	0.92	2.19	7.59	11.01	5.32
4.52	0.39	0.00	1.98	1.46	3.97	8.33	7.48	14.40
2.92	1.06	0.47	1.88	4.51	4.15	12.49	10.19	10.83
0.14	0.11	0.12	12.47	2.35	2.81	7.74	10.95	5.54
1.76	1.00	0.07	11.24	5.47	3.75	23.73	12.87	9.75
0.18	0.01	2.94	5.43	4.07	0.73	6.79	13.67	6.51
0.69	0.37	0.92	7.29	14.67	2.59	8.28	7.56	9.93
0.00	0.56	0.03	3.88	1.40	3.83	6.46	9.12	9.10
0.20	1.20	0.01	4.19	5.07	2.92	11.99	10.93	11.80
0.75	0.40	0.05	3.34	3.43	6.40	14.52	22.87	15.05
3.02	3.77	0.76	11.69	9.01	5.50	18.25	4.57	12.49
0.29	0.28	0.39	2.98	6.09	7.22	13.62	11.30	5.48
1.68	0.46	1.06	1.36	5.31	5.60	8.74	8.56	6.68
0.37	0.31	0.84	2.97	9.54	4.29	8.53	3.93	10.45
0.06	0.84	0.12	1.93	7.55	4.68	9.61	23.12	11.35
0.72	0.91	0.51	3.84	8.33	2.32	2.83	5.44	9.58
0.09	0.23	1.87	2.33	2.89	3.93	13.69	14.41	5.56
0.10	0.06	0.75	3.02	4.64	5.11	10.83	2.63	8.52
0.69	0.27	0.03	5.02	9.59	3.03	8.10	6.52	7.73
0.00	1.87	1.80	6.25	7.13	3.46	9.49	17.35	7.02
0.77	1.26	0.56	9.29	3.29	2.05	3.16	10.24	5.56
0.10	0.82	0.85	2.83	7.16	1.67	10.64	12.34	16.14
0.14	0.00	0.05	6.31	0.35	4.45	5.13	6.81	10.95
0.90	0.00	0.05	1.61	2.81	3.47	10.18	4.17	5.22
0.00	1.57	0.53	5.89	9.33	5.76	4.18	8.38	11.05
1.25	0.04	0.02	6.49	3.01	1.75	6.04	4.87	20.70
2.50	0.36	0.01	8.35	6.65	1.97	17.87	5.46	10.24
2.05	0.01	0.04	4.22	6.44	9.41	5.97	10.45	7.97
1.82	0.20		2.95	5.94		5.18	17.90	
1.76	0.00		2.61	5.43		10.19	3.44	

Obviamente la visualización de estos resultados no permite percibir las similitudes o diferencias entre las distribuciones muestreadas. Una forma de presentar estos resultados es el que se hace en la Figura 1.11. En ella se observan 3 cajas cuyo cuerpo está atravesado por una línea horizontal y de los extremos superior e inferior emergen sendos segmentos de recta que son continuados, en algunos casos, por cruces y círculos. La Figura 1.12 presenta esquemáticamente la interpretación de este gráfico.

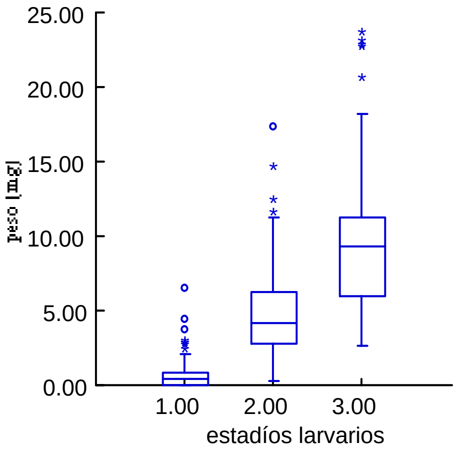


Figura 1.11: Diagramas de caja describiendo la distribución de pesos en tres estadios larvales.

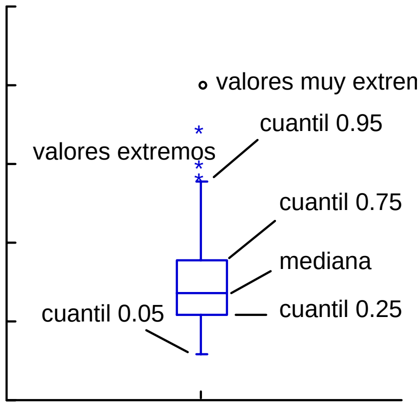


Figura 1.12: Diagrama explicativo de los objetos que aparecen en los diagramas de caja.

¿Qué se puede decir del peso de las larvas de los distintos estadios?

En primera instancia las distribuciones están posicionadas de manera diferente, siendo las larvas de estadio 3 las más pesadas, luego las de estadio 2 y finalmente las de estadio 1. Otro hecho que puede observarse es que la variación, al menos en términos absolutos, va incrementándose a medida que aumenta el peso promedio de las larvas. Un tercer hecho es que la distribución es asimétrica con valores extremos o muy extremos sólo a la derecha de la media. Por otra parte, la asimetría tiende a disminuir con el aumento del peso, esto indica que la distribución es más asimétrica en las larvas de estadio 1 que en las de estadio 2 ó 3. Esto se puede visualizar por la cantidad de valores muy extremos.

Diagrama de puntos (´Dot-Plot´)

Aunque el *box-plot* es una representación apropiada para la distribución de frecuencias muestrales, a veces el tamaño de la muestra es pequeño y los cuantiles muestrales que de ella se obtienen no son confiables desde el punto de vista estadístico y en consecuencia la construcción del *box-plot*, que requiere de estas medidas, puede no ser buena. En otras circunstancias no sólo se quiere tener una imagen de los aspectos generales de la distribución sino, también, una visualización de los valores efectivamente observados. En estos casos el *dot-plot*, puede ser la representación más satisfactoria. El procedimiento de construcción es simple y consiste en dibujar un punto por cada uno de los valores observados en la muestra, ubicados según una escala (la recta real) que se pone como referencia. Cuando hay más de una observación con el mismo valor, ésta se representa con otro punto ubicado en posición contigua al anterior y así sucesivamente con el resto de las observaciones repetidas.

Ejemplo 1.4

La siguiente tabla presenta los resultados observados del número de plántulas de malezas por m^2 en una muestra de tamaño n=20. La Figura 1.13, presenta el *dot-plot* para estos datos.

Tabla 1.4: Número de plántulas de maleza por m^2.

5	3	4	7	5
9	8	4	7	4
5	1	4	5	8
4	7	5	3	5

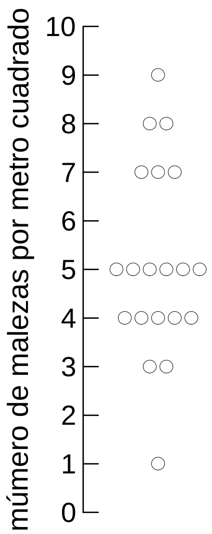


Figura 1.13: Diagrama de puntos para el número de malezas por m^2.

Histogramas y Polígonos

Otra forma alternativa de presentar los resultados de la Tabla 1.3 es mediante el clásico histograma. La Figura 1.14 presenta el histograma de frecuencias relativas y el polígono correspondiente al peso de las larvas del estadio 1. Lo más destacable que puede observarse es la marcada asimetría de la distribución; en comparación con la representación en *box-plot* es más difícil identificar los percentiles, los valores extremos y muy extremos.

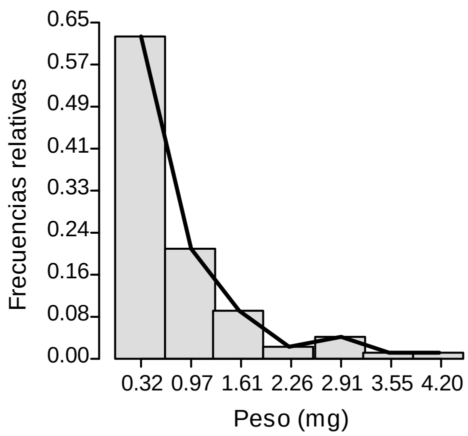


Figura 1.14: Histograma y polígono de frecuencias relativas del peso de larvas del estadio 1 de una especie de polilla forestal.

Diagramas de Tallo y Hojas

Otra forma de graficar la distribución de una variable es a través del diagrama de tallos y hojas. En la Figura 1.15 se muestra una representación del peso de las larvas de estadio 1 según un gráfico de tallos y hojas.

```
0      0000000000000000000000000000111111
0   M  222222222333333
0      444445555
0      66677777
0   H  88888999
1      000
1      222
1      45
1      677
1      8888
2      0
       *** Valores atípicos ***
2      5899
3      07
4      5
```

Figura 1.15: Diagrama de Tallos y Hojas para la variable peso de larvas de estadio 1 de una polilla forestal.

En esta figura el peso 0.47 (la primer larva de la Tabla 1.3), está representado en la tercer fila del gráfico mediante un número 4. Como hay cinco cuatros esto indica que hay cinco observaciones con cero como parte entera y cuatro como primer decimal. Para completar el ejemplo, y de acuerdo a la figura, el valor máximo tiene como entero al 4 y como primer decimal al 5. La fila marcada con "M" contiene al valor de la mediana y la marcada con "H" al punto a partir del cual los valores mayores (o menores si está por encima de la marca de la mediana) son extremos. La figura anterior muestra de manera muy clara el carácter asimétrico de la distribución de la variable estudiada y permite identificar con facilidad los valores atípicos.

Diagramas de Dispersión

Cuando se estudia la asociación entre 2 variables (por ejemplo X e Y) es muy útil hacer un diagrama de dispersión. Este es un gráfico en el que cada observación está representada en el plano XY por un punto cuyas coordenadas están dadas por los valores registrados en ambas variables. Por ejemplo, si se hace un experimento en maní en el que a distintas parcelas se agregan números crecientes de aplicaciones de un fungicida y se registra el rendimiento final, se podrían obtener los resultados mostrados en el Tabla 1.5. La Figura 1.16 presenta un diagrama de dispersión, donde se puede visualizar la existencia de una asociación positiva entre el rendimiento y el número de aplicaciones del fungicida. La Figura 1.17 sugiere una relación funcional curvilínea que liga al número de aplicaciones con el rendimiento obtenido.

Tabla 1.5: Rendimiento de maní para distinto número de aplicaciones de fungicida.

Aplicaciones	Rend.	Aplicaciones	Rend.	Aplicaciones	Rend.	Aplicaciones	Rend.
0	3.15	0	2.20	5	4.65	5	5.31
1	3.54	1	2.65	4	4.54	4	4.74
2	3.37	2	4.03	3	3.96	3	5.10
3	4.69	3	4.82	2	4.80	2	4.46
4	4.56	4	3.81	1	4.18	1	5.01
5	4.78	5	4.55	0	3.88	0	2.67
0	3.28	0	2.96	5	5.35	5	4.20
1	3.54	1	3.01	4	4.50	4	4.34
2	4.04	2	3.56	3	4.93	3	5.14
3	5.2	3	3.98	2	3.54	2	5.17
4	4.91	4	5.21	1	3.35	1	4.99
5	5.03	5	5.44	0	2.73	0	4.08

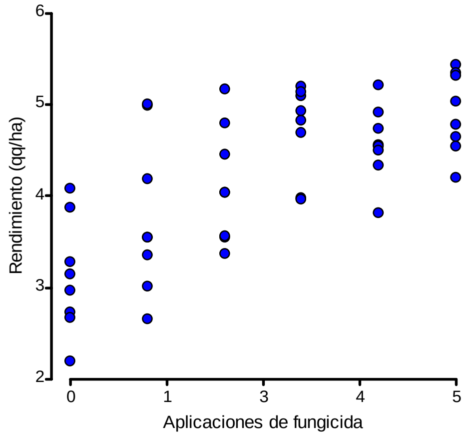


Figura 1.16: Diagrama de dispersión entre número de aplicaciones de fungicida y rendimiento de maní.

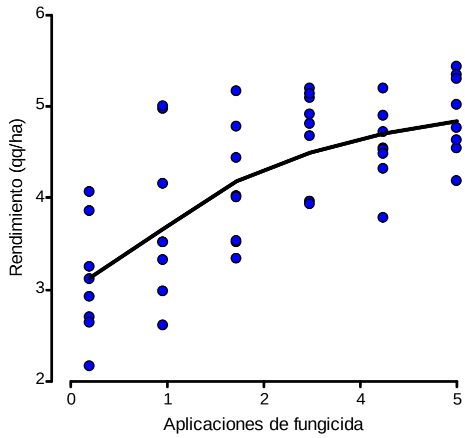


Figura 1.17: Diagrama de dispersión entre número de aplicaciones de fungicida y rendimiento de maní con una curva de ajuste que aproxima la relación entre estas variables.

Diagramas de Líneas

En algunos casos un diagrama de dispersión puede ser modificado incluyendo segmentos de recta que unen los puntos del plano según un orden dado por el eje de abscisas. Como ejemplo, supóngase que se evalúa el número de callos obtenidos en cultivos de 200 anteras sometidas a un número creciente de días de frío (Tabla 1.6.). La Figura 1.18 muestra estos resultados mediante un diagrama de líneas. Este permite visualizar con claridad la tendencia decreciente del número de callos formados en función del número de días de frío y la forma en que esto ocurre. Así, se observa una fuerte caída inicial para luego llegar a una situación de estabilidad con una leve tendencia decreciente.

Tabla 1.6: Días de frío y número de callos obtenidos a partir de 200 anteras cultivadas.

Días	0	1	2	3	4	5	6
# Callos	150	50	45	40	38	27	15

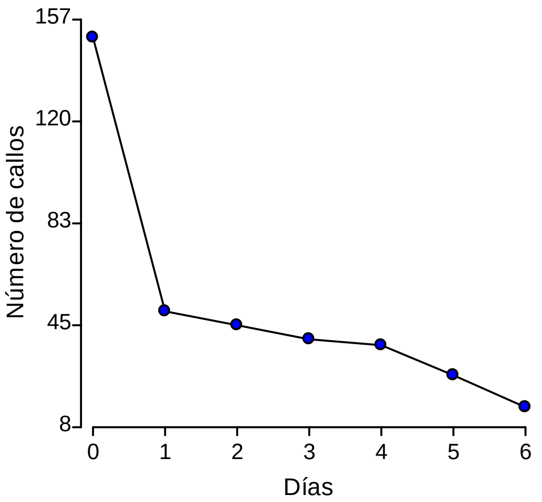


Figura 1.18: Diagrama de líneas que muestra la relación entre días de frío y número de callos formados sobre 200 anteras cultivadas.

'Q-Q Plots'

Este gráfico permite la comparación de la distribución de frecuencias de una variable con una distribución teórica. Una de estas distribuciones, que se estudiará más adelante, es la distribución normal. En ese caso se habla de 'Q-Q plot' normal. El nombre proviene del hecho de representar en él los cuantiles muestrales versus los cuantiles teóricos (*quantil to quantil plot*). Así, si se presupone que la distribución de la variable altura de hipocótilo de una especie de Prosopis es una variable normal, se podría verificar esto gráficamente mediante un 'Q-Q plot'. Este gráfico no es más que un diagrama de dispersión donde los valores de los ejes X e Y se obtienen según el siguiente algoritmo.

a) Ordenar la muestra de menor a mayor y designar al valor con la posición i-ésima como $x^{[i]}$. Sean $\overline{x}$ y S, la media y la desviación estándar muestrales correspondientes,

b) Para cada observación ordenada obtener las coordenadas (X,Y) para construir el gráfico 'Q-Q plot' siendo:

- $X_i = x^{[i]}$ (el primer elemento de X es el dato más pequeño de la muestra, $x^{[1]}$, y el último elemento es el mayor valor observado, $x^{[n]}$).

- $Y_i = \Phi^{-1}((i\text{-}0.5)/n) \cdot S + \overline{x}$, donde Φ^{-1} es la función inversa de la función de distribución normal estándar. Los resultados de esta función se obtienen de una tabla de distribución normal (Tabla Normal del Apéndice) buscando el argumento en la columna que dice $P(Z \leq z)$ y como resultado el valor de z correspondiente.

Por ejemplo, la siguiente tabla muestra los valores observados de longitud del folíolo en 30 hojas de garbanzo.

Tabla 1.7: Longitud del folíolo en plantas de garbanzo.

3.0	2.7	2.9	3.2	3.0	3.5	3.4	2.8	3.3	2.9
3.1	2.3	2.9	3.1	3.4	2.9	3.3	3.0	2.7	3.1
2.8	2.8	3.1	2.5	3.1	2.8	2.8	3.1	2.5	3.2

Ordenando los datos, los correspondientes valores de X e Y del 'Q-Q plot' se muestran a continuación:

X	(i-0.5/n)	Y	X	(i-0.5/n)	Y	X	(i-0.5/n)	Y
2.3	0.017	2.39	2.9	0.350	2.87	3.1	0.683	3.11
2.5	0.050	2.52	2.9	0.383	2.89	3.1	0.717	3.13
2.5	0.083	2.59	2.9	0.417	2.92	3.1	0.750	3.16
2.7	0.117	2.65	2.9	0.450	2.94	3.2	0.783	3.19
2.7	0.150	2.69	3.0	0.483	2.96	3.2	0.817	3.22
2.8	0.183	2.73	3.0	0.517	2.99	3.3	0.850	3.26
2.8	0.217	2.76	3.0	0.550	3.01	3.3	0.883	3.30
2.8	0.250	2.79	3.1	0.583	3.03	3.4	0.917	3.36
2.8	0.283	2.82	3.1	0.617	3.06	3.4	0.950	3.43
2.8	0.317	2.84	3.1	0.650	3.08	3.5	0.983	3.56

El gráfico resultante se muestra en la Figura 1.19. Cuando la distribución de la variable coincide con la del modelo propuesto, entonces los puntos X,Y se alinean en una recta a 45° (pendiente 1), como en el caso presentado. Una variante del 'Q-Q plot' es el 'P-P plot' que grafica percentiles vs. percentiles.

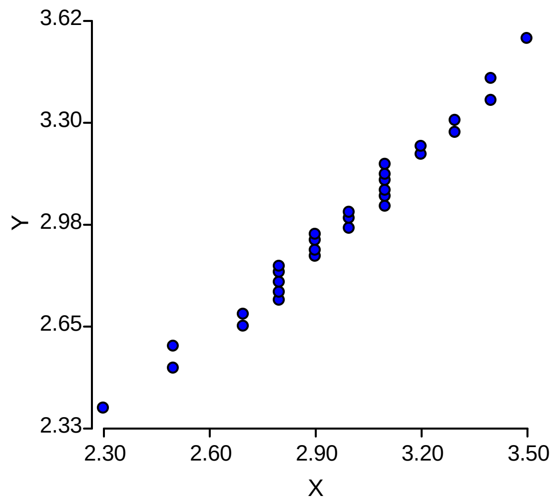


Figura 1.19: Q-Q plot normal para los datos presentados en la Tabla 1.7.

Ejercicios

Ejercicio 1.1

A partir de las situaciones que se describen a continuación, identificar la población en estudio y una o más variables que sean de utilidad para el análisis del problema en cuestión.

Situación A: En una zona del departamento de Río Primero, en la Provincia de Córdoba, donde se cultiva zapallo para obtención de semillas, se observó que las cosechas de semillas del último trienio disminuyeron considerablemente con respecto a períodos anteriores, aún cuando el área cultivada se mantenía sin cambios. Entrevistas con técnicos de la zona revelaron que varias podrían ser las causas de tal disminución en los rendimientos. Entre ellas se consideraban especialmente:

1) Una infestación varietal producida por el cruzamiento de las poblaciones para cosecha, con las poblaciones de zapallito amargo, que enmalezan los cultivos de maíz de la zona. Se conoce por investigaciones previas que cuando se produce este tipo de hibridación los zapallos cultivados dan flores con distinto número de pétalos y disminuyen la producción de semillas.
2) Un aborto de óvulos, generadores de semillas, por influencia de las pulverizaciones que se han introducido en los últimos tres años. El efecto visible de las pulverizaciones es el amarilleo y la disminución del diámetro de los

ovarios.

Situación B: En una experiencia de laboratorio se ha inoculado un complejo virósico a trescientas macetas que contienen plántulas de tabaco. Se cree que dicho complejo puede provocar diversos grados de clorosis en el follaje o bien no producir clorosis, pero disminuir considerablemente la altura de plántulas.

Ejercicio 1.2

Basándose en una situación problemática que sea de interés, plantear un experimento y un estudio observacional. Definir para cada uno la/s variable/s respuesta/s.

Ejercicio 1.3

Clasificar las siguientes variables en continuas o discretas:

a) Número de semillas de alfalfa por metro de surco sembrado.
b) Temperaturas registradas cada media hora en un laboratorio, durante una semana.
c) Período de tiempo desde el almacenamiento y hasta que se produce el deterioro del 50% de los frutos almacenados.
d) Milímetros de precipitación de una localidad durante un año.
e) Número de semillas en dormición en cajas de 50 semillas.
f) Número de materias aprobadas con 4 puntos por estudiantes de la Facultad de Agronomía durante el período 1994-1997.
g) Cociente entre el largo y el ancho de los entrenudos de plantas de maíz.

Ejercicio 1.4 (Para hacer en el aula)

MUESTRAS ALEATORIAS VERSUS MUESTRAS APLICANDO SU JUICIO

¿Cuál es mejor?

Supóngase que una persona está interesada en conocer cuál es la superficie promedio de los lotes de una región. Para ello debe seleccionar entre los siguientes métodos:

Método 1: extraer una muestra de lotes que considere “representativa” o buena a su juicio, y calcular el promedio de la misma.

Método 2: extraer una muestra aleatoria y calcular el promedio de la muestra.

Para analizar las consecuencias de la selección de uno u otro método y del tamaño de la muestra, realizar el siguiente experimento:

Método 1: muestra aplicando su juicio

a) Mirar durante 10 segundos la hoja con la figura adjunta a este ejercicio y arriesgar una cifra para el promedio del área de los rectángulos en la página. La unidad de medida es el cuadrado unitario; por ejemplo un rectángulo de 3 filas por 4 columnas de cuadraditos tiene un área de 12. Tal esquema podría representar un lote de 12 hectáreas. Anotar el resultado de la inspección visual.

b) Obtener las muestras:

   1) *Primera muestra*: seleccionar 5 rectángulos o lotes, que a su juicio, sean representativos de los rectángulos en la página. Anotar el número de cada uno de los 5 lotes, el cual se encuentra al pie de cada uno de ellos. Anotar las áreas de cada uno de estos lotes, después calcular el promedio de las 5 áreas.

   2) *Segunda muestra*: repetir la parte 1) pero seleccionando 15 rectángulos. Registrar el promedio de las 15 áreas.

Recoger todos los valores obtenidos en la clase de la partes a), y b). Hacer un gráfico para cada uno de los tres conjuntos de valores a los fines de observar alguna tendencia.

Método 2: muestra aleatoria

a) Usando los números de los rectángulos y la tabla de números aleatorios, seleccionar 5 rectángulos aleatoriamente. Escribir los números y sus correspondientes áreas, y luego calcular el promedio de estas.

b) Repetir lo realizado en el punto anterior para un conjunto de 15 rectángulos.

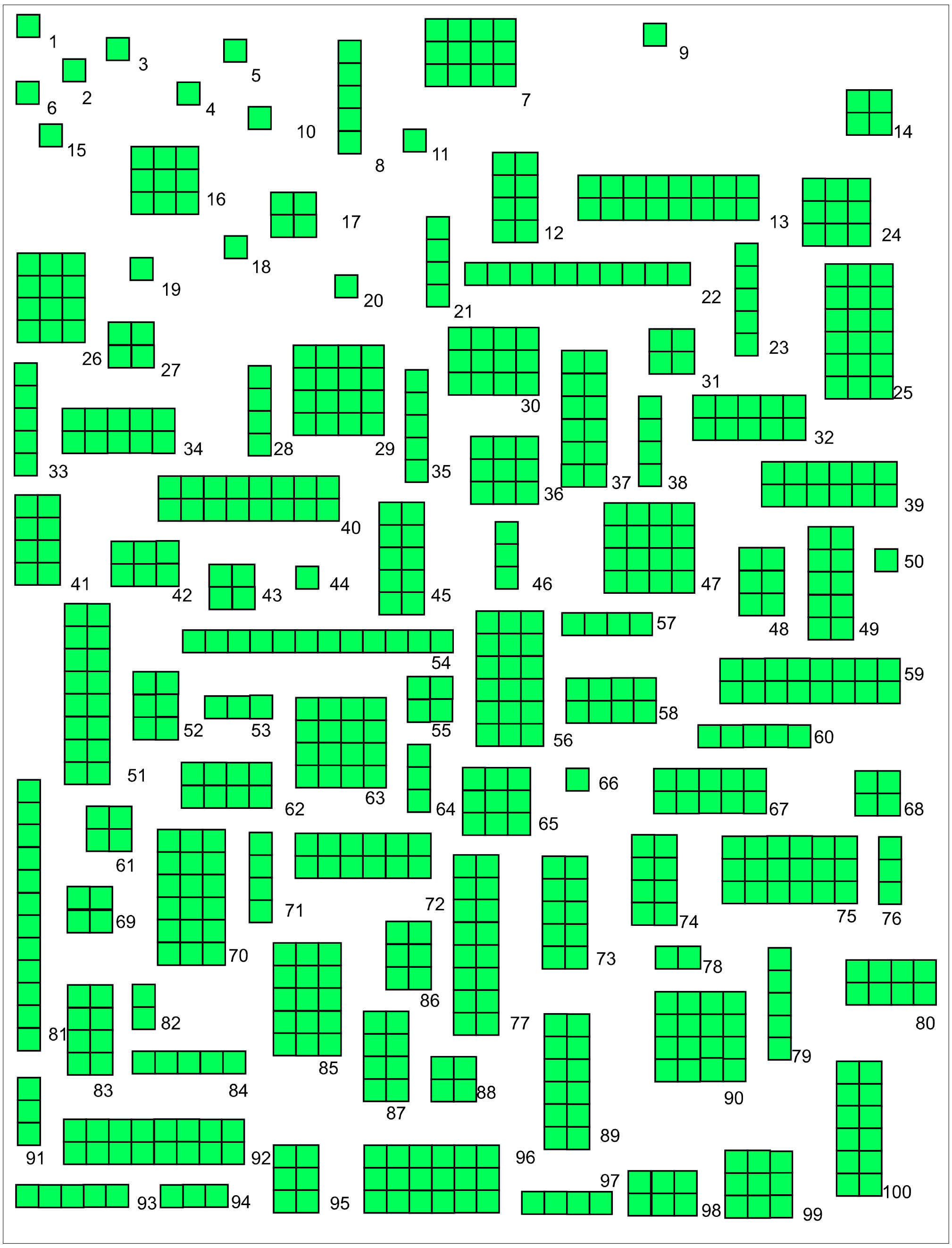
c) Calcular el promedio de las 20 áreas de los ítems a) y b).

Hacer los gráficos con los promedios obtenidos por cada uno de los alumnos en los ítems a), b) y c) y compararlos con los obtenidos en el método 1.

Teniendo en cuenta que la media poblacional de este conjunto de lotes es 7.5 hectáreas, responder las siguientes preguntas:

a) Muestra aleatoria versus muestra aplicando su juicio. ¿Cuál produce menor sesgo?

b) Dadas las estimaciones con $n = 5$, $n = 15$ y $n = 20$, ¿Cuál es más precisa?



Ejercicio 1.5

La siguiente tabla muestra la distribución de frecuencias de la *variable salarios mensuales* (en pesos), obtenida en un muestreo aleatorio de 65 empleados de una firma agropecuaria:

Salario	***Nº de Empleados***
(500 - 600]	8
(600 - 700]	10
(700 - 800]	16
(800 - 900]	14
(900 - 1000]	10
(1000 - 1100]	5
(1100 - 1200)	2

a) Representar gráficamente la distribución de frecuencias de la variable.

b) ¿Qué porcentaje de empleados tiene salario inferior o igual a 800 pesos? ¿Es éste el cuantil 0.80?

c) ¿Qué porcentaje de empleados tiene salario mayor a 800 pesos?

d) Calcular los cuantiles 0.50 y 0.30 de la distribución.

Ejercicio 1.6

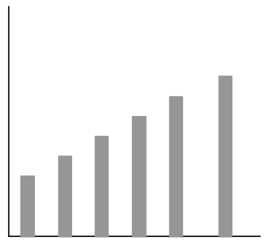
A partir de la observación de los siguientes gráficos, ¿qué diagrama se asocia con cada una de las siguientes descripciones?

a) Distribución de la población argentina en 1990 según la edad (en años). El rango es de 0 a 90, el tamaño de la clase o amplitud del intervalo es 10.

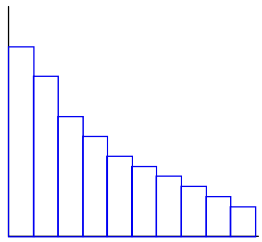
b) Distribución del número de plantas muertas con relación a la severidad de una enfermedad. La severidad se mide de acuerdo a una escala categórica de 0 a 5 en orden creciente de ataque.

c) Distribución de altura de plantas en un cultivo de trigo (en cm.). Rango de 0 a 50, tamaño de clase 5.

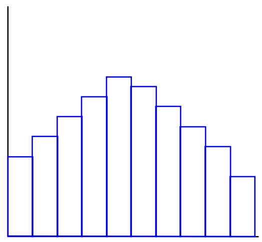
d) Distribución de personas según la distancia (en Km.) que transitan desde su hogar al trabajo. El rango va de 0 a 50, el tamaño de clase es 5.



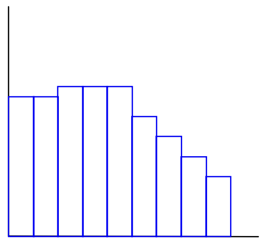
1º



2º



3º



4º

Ejercicio 1.7

Dentro de las actividades agrícolas del Departamento Tulumba de la Provincia de Córdoba, durante el período 88-89, predominó el cultivo de maíz con 60000 tn. producidas, mientras que de soja se obtuvieron 3000 tn. y de sorgo 2000 tn.

Representar gráficamente el comportamiento de la variable producción para cada cultivo.

Ejercicio 1.8

Los siguientes datos corresponden a la ganancia de peso por día (expresada en gramos), de novillos sometidos a una dieta experimental.

704	890	986	806	798	995	876	705	706	915
801	720	807	960	858	606	798	708	893	906
660	780	615	895	969	880	700	697	804	918
825	809	758	705	800	910	896	708	690	830

Obtener medidas descriptivas, graficar e interpretar la información contenida en esta muestra.

Ejercicio 1.9

Los siguientes datos se refieren al número de dientes por hoja en bulbos de ajo:

4	2	2	3	3	2	3	3	2	2
3	3	2	1	2	2	2	2	4	2
4	2	3	3	1					

a) Construir la tabla de distribución de frecuencias y representarla gráficamente.

b) ¿Cuál es la proporción o probabilidad aproximada de encontrar hojas con menos de 2 dientes?

c) ¿Cuál es la proporción o probabilidad aproximada de encontrar hojas con más de 2 dientes?

Ejercicio 1.10

En un estudio en un monte del Chaco árido se midieron los perímetros basales de troncos de plantas de quebracho blanco (en centímetros) y se obtuvo la siguiente información.

138	164	150	132	144	125	149	157	146	158
140	147	136	148	152	144	168	126	138	176
163	119	154	165	146	173	142	147	135	153
140	135	161	145	135	161	145	142	150	156
145	128								

a) Construir la tabla de distribución de frecuencias y representarla gráficamente.

b) Obtener las siguientes medidas: media, mediana, modo, $X_{0.25}$, $X_{0.75}$, rango, desviación estándar y coeficiente de variación.

Ejercicio 1.11

Una compañía dedicada a la comercialización de semillas decidió poner a prueba el rendimiento de dos híbridos de sorgo granífero bajo riego. Se estudiaron dos muestras, una del híbrido "Nueva GR80" y otra del híbrido "Overa". Los resultados, en qq/ha fueron:

Nueva GR80:

110 112 135 140 128 132 123 125 140 142 151 113 142 123 118 143 138 135 140 135

112 128 152 136 152 139 142 129 150 135 119 128 123 142 138 145 136 147 141 137

Overa:

115 158 139 143 151 152 148 139 153 125 136 129 146 136 140 150 140 139 128 129

125 130 140 149 150 139 142 138 129 126 137 148 146 150 158 153 119 139 154 139

151 154 139 132

a) En base a las medidas muestrales, ¿cuál de los dos híbridos recomendaría?

b) Representar gráficamente ambas muestras.

2

Variables Aleatorias

Introducción

En este capítulo se darán las definiciones de **espacio muestral**, **punto muestral**, **evento**, y **evento aleatorio,** necesarias para introducir el **concepto de variable aleatoria**.

Se darán además tres definiciones de probabilidad, la de **Kolmogorov**, la **frecuencial** y la **clásica**. También se presentará una definición de **función de densidad y distribución,** que abarca tanto el caso de variables aleatorias continuas como discretas. Finalmente se introducirá el concepto de **parámetros** para la caracterización de distribuciones de variables aleatorias.

Espacio Muestral - Eventos

Ejemplo 2.1

Considérese un experimento que consta de la observación de 3 semillas en un cierto orden, cada una de las cuales puede estar **sana** (situación que se representará con el signo "**+**") o bien **enferma** (situación que se representará con el signo "**-**"). Hay 8 resultados posibles en el experimento, los que conforman un conjunto que se denomina *espacio muestral* y que a continuación se representa:

$$\Omega = \{+++\,,\,+--\,,\,++-\,,\,--+\,,\,+-+\,,\,-+-\,,\,-++\,,\,---\}$$

Definición 2.1: Espacio muestral

Se llama **espacio muestral** al conjunto de todos los resultados posibles de un proceso experimental u observacional. Será denotado con la letra griega omega (Ω).

Definición 2.2: Punto muestral o evento elemental

Se llama **punto muestral o evento elemental** a cada uno de los elementos del conjunto Ω y será denotado genéricamente como: ω.

Siguiendo con el ejemplo, un punto muestral es el resultado posible "*tres semillas sanas"* (representado por ω = (+ + +)), otro punto muestral es "*la primera semilla sana y las otras dos no"* (ω = (+ - -)).

Definición 2.3: Evento

Dado un espacio muestral Ω, se llama **evento** a cualquier subconjunto de Ω.

Se utilizarán letras A, B, C,.... para denotar los eventos, salvo en el caso del mismo Ω (llamado *evento cierto*) y del subconjunto que no contiene ningún resultado posible (conjunto vacío), denotado por ∅.

Un evento de Ω, puede ser "observar una semilla cualquiera sana y las otras no". Este evento esta constituido por los siguientes puntos muestrales:

$$A = \{+ - - , - + - , - - +\}.$$

Definición 2.4: Eventos mutuamente excluyentes

Se dice que dos eventos A y B de un espacio muestral Ω son **mutuamente excluyentes** si no contienen elementos en común, o sea si la intersección de A y B es el conjunto vacío ($A \cap B = \emptyset$).

Ejemplo 2.2

Dados A = {+ - - , - + - , - - +} y B = {+ + - , - + + , + - +}, entonces puede observarse claramente que A y B son eventos mutuamente excluyentes.

Por la teoría de conjuntos se tiene:

Si A y B son dos eventos de Ω, la **unión** de estos eventos conforma un nuevo conjunto, que contiene a los puntos muestrales de A y de B. La unión de A y B se denota por $A \cup B$.

Si A y B son dos eventos de Ω, la **intersección** de estos eventos conforma un nuevo conjunto, que contiene a los puntos muestrales que pertenecen a A y a B simultáneamente. Se denota la intersección de A y B por $A \cap B$.

Ejemplo 2.3

Sean: A = {+ - - , - + - , - - +}   y   B = {+ + - , + - + , - + +}.
Entonces, $A \cup B$ = {+ - - , - + - , - - + , + + - , + - + , - + +}

Ejemplo 2.4

Sean A el evento “observar exactamente una semilla sana” y el evento B “observar que la segunda semilla esté sana” (B={- + -, + + +, + + -, - + +}), luego $A \cap B$ ={- + -}.

Ejemplo 2.5

Sean los eventos A ={+ + +} y B ={- - -}, luego, puede verse fácilmente que $A \cap B$ resulta ser el conjunto vacío.

Probabilidad

Se analizará primero la definición de probabilidad dada por Kolmogorov y luego las definiciones frecuencial y clásica.

Probabilidad según Kolmogorov

Esta definición es la más amplia y la menos intuitiva, pero sienta las bases para el desarrollo de la teoría de probabilidades, en tanto las otras dos definiciones brindan criterios de cálculo de la probabilidad de un evento.

Definición 2.5: Medida de Probabilidad (Kolmogorov, 1937)

Sea Ω un espacio muestral. La función P(.) que asigna a cada evento de Ω un número real en el intervalo [0,1], se llama medida de probabilidad si satisface los siguientes axiomas:

i. $P(\Omega) = 1$

ii. $P(A) \geq 0$, donde A representa un evento cualquiera de Ω

iii. Si A_1, A_2, ... es una secuencia de eventos mutuamente excluyentes entonces:
$P\left(U_i A_i\right) = \sum_i P\left(A_i\right)$.

De esta definición se deduce que dados los eventos A y B, la probabilidad de que ocurra A ó B es dada por $P(A \cup B) = P(A) + P(B) - P(A \cap B)$, donde $P(A \cap B)$ denota la probabilidad de que ocurran A y B simultáneamente.

Si A y B son mutuamente excluyentes, $A \cap B = \emptyset \Rightarrow P(A \cap B) = P(\emptyset) = 0$. En otro caso, $0<P(A\cap B) \leq 1$.

Definición 2.6: Probabilidad condicional

Para dos eventos A y B que pertenecen al mismo espacio muestral la probabilidad condicional se define como:

$$P(A|B) = P(A\cap B)/P(B),$$

donde $P(A|B)$ denota la probabilidad condicional de ocurrencia del evento A dado que ha ocurrido B.

Por ejemplo, si B es el evento "día nublado" y A el evento "llueve en el día", luego $P(A|B)$ denotará la probabilidad de lluvia dado que el día está nublado.

Definición 2.7: Independencia de Eventos

Dados A y B eventos que pertenecen al mismo espacio muestral, se dice que son independientes si $P(A\cap B) = P(A).P(B)$.

Luego, si A y B son eventos independientes se tiene que $P(A|B).P(B) = P(A).P(B)$.

Ejemplo 2.6

Considérese un experimento en el que se analizan 3 pariciones de una vaca (n=3), registrándose el sexo del ternero nacido. Como los resultados posibles de cada parición son dos (N=2), los resultados posibles del experimento son $N^n=2^3=8$. Estos son: HHH, HHM, HMH, MHH, HMM, MHM, MMH y MMM donde M representa una cría macho y H una cría hembra y se asume que estos resultados son igualmente probables. Defina los eventos A como "una cría hembra nace en cada uno de los dos primeros partos"; B como "un macho nace en el tercer parto" y C como "exactamente 2 machos ocurren en los tres partos". Mostrar que A y B son dos eventos independientes, mientras que B y C no lo son:

$$A = \{HHH , HHM\} \qquad B = \{HHM , HMM , MHM , MMM\}$$

$$C = \{HMM , MHM , MMH\}$$

A y B son independientes $\Leftrightarrow P(A\cap B) = P(A) . P(B)$

Como $A \cap B$ = {HHM} $\Rightarrow P(A \cap B) = \frac{1}{8}$. Por otra parte, $P(A) = \frac{2}{8}$ y $P(B) = \frac{4}{8}$ y $P(A).P(B) = \frac{2}{8} \cdot \frac{4}{8} = \frac{1}{8}$. Luego, $P(A \cap B) = P(A) .P(B)$, por consiguiente A y B **<u>son</u> independientes**.

Si B y C son independientes $\Rightarrow P(B \cap C) = P(B) . P(C)$. En este ejemplo, $P(B) = \frac{4}{8}$ y $P(C) = \frac{3}{8}$. Así, $P(B) . P(C) = \frac{4}{8} . \frac{3}{8} = \frac{3}{16}$. Por otro lado, $B \cap C$ = {HMM , MHM} $\Rightarrow P(B \cap C) = \frac{2}{8} \neq P(B) . P(C)$. Luego B y C **<u>no son</u> independientes**.

Probabilidad: Concepto Frecuencial

Definición 2.8: *Probabilidad: concepto frecuencial*

Si A es un evento y n_A es el número de veces que A ocurre en N repeticiones independientes del experimento, la **probabilidad del evento A**, denotada por P(A), se define como:

$$\lim_{N \to \infty} \frac{n_A}{N} .$$

Observación 1: Un estudio o ensayo es independiente de otro, si los resultados de uno no influyen sobre los resultados del otro. Cuando el resultado de un ensayo brinda información sobre la probabilidad de obtener un resultado determinado en otro ensayo se dice que ambos ensayos no son independientes.

Observación 2: La noción de límite para $N \to \infty$ debe ser interpretada para "N suficientemente grande".

Ejemplo 2.7

Considérese que la observación de una semilla es un ensayo. Suponga que con A se representa el evento "encontrar la semilla germinada". Si se observan 1000 semillas (se repite 1000 veces el ensayo, N = 1000), en condiciones tales que cada observación sea independiente una de otra[2] y si 600 semillas germinan (n_A = 600), se dice que la probabilidad estimada de observar una semilla germinada, está dada por:

2 Los ensayos deberían planificarse de manera tal que el hecho de que una semilla germine o no, no determine o afecte la probabilidad de germinación de las otras semillas.

$$P(A) = P(\text{observar una semilla germinada}) = \frac{n_A}{N} = 600 / 1000 = 0.6$$

Observación 3: En este caso se habla de probabilidad estimada o aproximada por una cierta proporción ya que se usó la noción de límite para calcular P(A).

Probabilidad: Concepto Clásico

Cuando Ω es finito (el número de puntos muestrales es finito) se puede dar otra definición de probabilidad, que es la que se desarrolló originariamente estudiando los juegos de azar.

Definición 2.9: Probabilidad: concepto clásico

Dado el evento A en Ω,

$$P(A) = \frac{\text{Número de puntos muestrales favorables}}{\text{Número total de puntos muestrales}}$$

Ejemplo 2.8

Supongamos que el experimento consiste en arrojar un dado una única vez. El espacio muestral asociado es Ω = {1, 2, 3, 4, 5, 6 }. Si se asigna a cada punto muestral la misma probabilidad de ocurrencia, es decir P(ω) = 1/6 para todo ω ∈ Ω; y si el evento de interés (A) es "que salga un número par", el cual consta de los puntos muestrales {2, 4, 6}, entonces, se tiene que P(A) = 3/6 = 0.5.

Evento Aleatorio

Definición 2.10: Evento aleatorio

Un evento A al cual atribuimos una probabilidad será llamado **evento aleatorio**.

Concepto de Variable Aleatoria

Una variable aleatoria es una función que asocia a cada elemento del espacio muestral Ω un número real. En el esquema que se presenta a continuación (Figura 2.1), se observa un espacio muestral Ω, conformado por N puntos muestrales $\{\varpi_i$: i=1,...,N$\}$ y un evento aleatorio A conformado por dos puntos muestrales, $\{\varpi_1, \varpi_N\}$. Un ejemplo de variable aleatoria, X, puede ser la función que toma a ϖ_1 (lo que se denota por $X(\varpi_1)$) y le asocia el número 154, y a ϖ_N le asigna el 111, esto es $X(\varpi_N) = 111$.

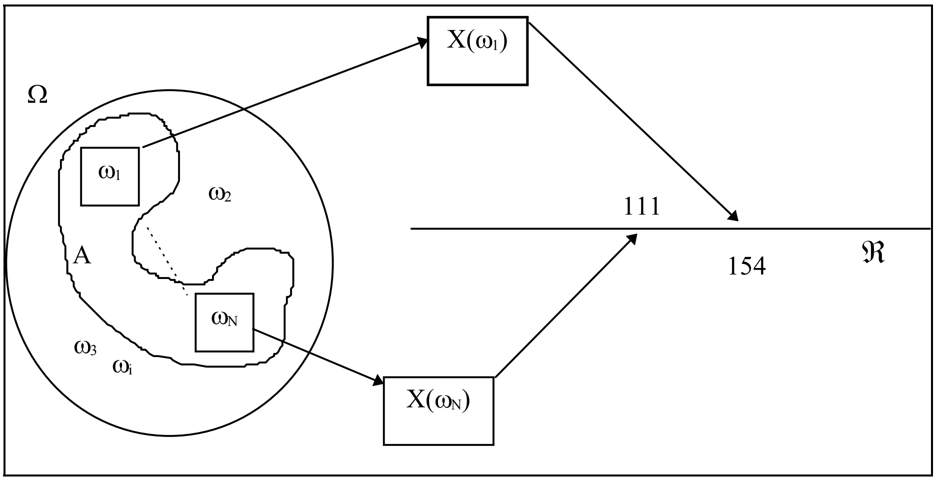


Figura 2.1 Representación del espacio muestral, eventos elementales, eventos y variable aleatoria

Para introducir el concepto de variable aleatoria es necesario presentar algunos eventos particulares, como son:

$[X \le a] = \{\varpi \in \Omega : X(\varpi) \le a\}$
$[X = a] = \{\varpi \in \Omega : X(\varpi) = a\}$

donde Ω es el espacio muestral que los contiene. Obsérvese que para un valor arbitrario *a* dado, se puede identificar un conjunto de Ω tal que $X(\varpi) \le a$.

Definición 2.11: Variable aleatoria

Dado un espacio muestral Ω con un probabilidad asociada, una **variable aleatoria** X es una función real definida en Ω tal que $[X \le x]$ es un evento aleatorio $\forall\ x \in \Re$. O sea $X: \Omega \rightarrow B \subseteq \Re$ es una variable aleatoria si para cualquier $x \in \Re$, $[X \le x]$ es un evento aleatorio.

Si el conjunto B, contenido en $\Re$ es no numerable, la variable aleatoria será llamada **variable aleatoria continua**. Si B es un conjunto numerable la variable será llamada **variable aleatoria discreta**.

Ejemplo 2.9

Supóngase que el experimento consiste en arrojar una vez una moneda y observar si salió cara o cruz. En este caso, Ω = {cara, cruz}. Si se asigna la misma probabilidad a cada punto muestral y se define la variable X como el número de caras, se tiene que X(cara) =1 y X(cruz) = 0, o sea X es una variable aleatoria discreta. Los eventos que se pueden definir en este espacio muestral son: {cara}, {cruz}, Ω, ∅. Obsérvese que X satisface la definición de variable aleatoria: Si $x < 0$, el conjunto $\{\omega:X(\omega) \leq x\} = \emptyset$. Si $0 \leq x < 1$, el conjunto $\{\omega:X(\omega) \leq x\} = \{cruz\}$. Si $x \geq 1$, el conjunto $\{\omega:X(\omega) \leq x\} = \Omega$. Luego, los eventos $[X \leq x]$ para cualquier x, tienen asociado una probabilidad, se puede decir entonces que X es una variable aleatoria.

Ejemplo 2.10

Considérese el experimento dado en el Ejemplo 2.1. Recordemos que el experimento consistía en la observación de 3 semillas en un cierto orden, cada una de las cuales puede estar sana (situación que se representó con el signo "+") o bien enferma (con el signo "-"). El espacio muestral de este experimento es:

$$\Omega = \{+ + + , + + - , + - + , - + + ,\ + - - , - - + , - + - , - - -\}$$

Definamos ahora la variable aleatoria X como el *número de semillas sanas*; luego, asumiendo igual probabilidad para cada punto muestral, tenemos que:

$$X (+ + +) = 3 \text{ y } P (X = 3) = 1/8$$

$$X (- - -) = 0 \text{ y } P (X = 0) = 1/8$$

$$X (+ - -) = X (- + -) = X (- - +) = 1 \text{ y } P(X = 1) = 3/8$$

Esto último se deduce del axioma iii de la definición de probabilidad y por ser estos eventos excluyentes. La Figura 2.2 representa esquemáticamente la variable aleatoria en cuestión.

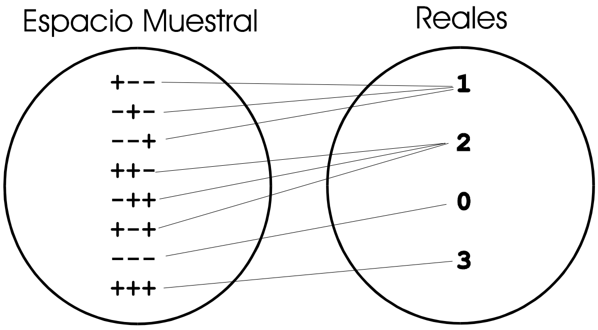


Figura 2.2: Representación de la variable aleatoria "número de semillas sanas" aplicada al espacio generado por el experimento del Ejemplo 2.10.

Distribución de una Variable Aleatoria

Como se presentó en las secciones anteriores, una variable aleatoria es un descriptor de eventos aleatorios y su función de distribución asigna probabilidades a esos eventos.

Función de Distribución Acumulada

A continuación se da una definición de función de distribución acumulada que abarca tanto el caso de variables aleatorias continuas como discretas.

Definición 2.12: Función de distribución acumulada

La **función de distribución acumulada**, o simplemente función de distribución, de una variable aleatoria X, denotada por F(.), es una función $F:\Re\rightarrow[0,1]$ tal que:

$$F(x) = P([X \leq x]) \ \forall \ x \in \Re.$$

¿Cómo debe leerse $F(x) = P([X \leq x]) \ \forall \ x \in \Re$? Como se recordará, $A = [X \leq x]$ describe un evento en particular, esto es, aquel conjunto de puntos muestrales a los que la variable aleatoria les asocia un valor menor o igual que x.

Luego $P([X \leq x])$ denota la probabilidad del evento $A=[X \leq x]$; $F(x) = P(A)$ indica que se asigna a F(x) el valor de probabilidad del evento $[X \leq x]$ ya que esto se cumple para todo x que pertenece al conjunto de números reales.

En otras palabras, la definición dice que "la función F evaluada en el punto x es la probabilidad de que la variable aleatoria X tome un valor menor o igual al punto x, para todo x que pertenece a los reales".

Ejemplo 2.11

Siguiendo con el Ejemplo 2.1 y según con lo planteado en el Ejemplo 2.10, ¿cuál es la función de distribución acumulada de la variable número de semillas sanas? Para contestar esta pregunta se puede realizar una tabla de dos columnas. En la primera de ellas, se detallan los posibles valores de la variable X y en la segunda F(x).

x	F(x)
0	F(0) = P(0) = 1/8
1	F(1) = P(0) + P(1) = 1/8 + 3/8 = 4/8
2	F(2) = P(0) + P(1) + P(2) = 1/8 + 3/8 + 3/8 = 7/8
3	F(3) = P(0) + P(1) + P(2) + P(3) = 8/8 = 1

Así, se tiene que:

a) F(x) = 0 para valores de x < 0

b) F(x) = 1/8 para 0 ≤ x < 1

c) F(x) = 1/2 para 1 ≤ x < 2

d) F(x) = 7/8 para 2 ≤ x < 3

e) F(x) = 1 para x ≥ 3

El gráfico de esta función de distribución acumulada será:

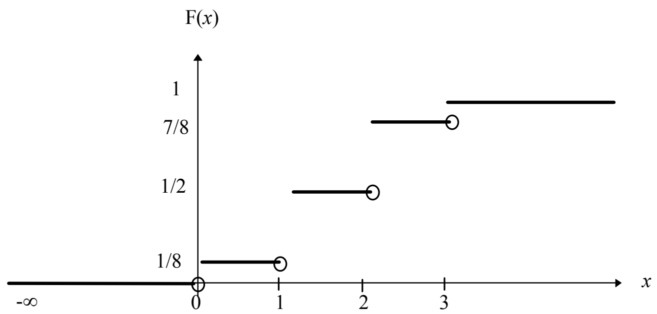


Figura 2.3: Gráfico de la función de distribución de la variable aleatoria "número de semillas sanas" en un experimento en el que se observan 3 semillas y donde todos los puntos muestrales son equiprobables.

Acerca del gráfico hay que destacar los siguientes aspectos:

a) Aunque la variable sea discreta, la F(x) esta definida para todo x en los reales, por eso se representa desde -∞ hasta ∞.

b) En los valores de x que coinciden con los valores que puede asumir la variable aleatoria se produce un salto que es igual a P(x).

c) El círculo que delimita el extremo de los segmentos, denota que en ese punto F(x) "salta".

Función de Densidad

Se ha visto que la función de distribución acumulada de la variable aleatoria X asocia a cada número real x la probabilidad de [X ≤ x]. Pero también se puede preguntar ¿cuál es la probabilidad de que X = x? Para responder a esta pregunta se distinguen dos casos: el de las variables aleatorias discretas y el de las variables aleatorias continuas.

La respuesta se deriva a partir de una función llamada función de densidad o simplemente densidad de la variable aleatoria. En el caso de las variables aleatorias continuas la respuesta es siempre la misma P(X = x) = 0, en cambio, para las variables discretas, P(X = x) ≥ 0.

Función de densidad de una variable aleatoria discreta

Definición 2.13: Función de densidad de una v.a. discreta

La **función de densidad** de una variable aleatoria discreta, denotada por f(.), es una función $f{:}\Re \to [0,1]$ tal que:

$$f(x) = \begin{cases} P(X = x) \text{si} & x \in C \\ 0 \text{ en caso contrario} \end{cases}$$

donde C = { x_1, x_2, x_3, ... } es el conjunto de valores que puede tomar la variable aleatoria discreta.

Retomando el concepto de función de distribución acumulada, notemos que cuando X es una variable aleatoria discreta, F(.) puede ser definida a partir de la función de densidad discreta de la siguiente manera:

$$F(X) = \sum_{x_i \le x \wedge x_i \in C} P(X = x_i) = \sum_{x_i \le x \wedge x_i \in C} f(x_i)$$

Ejemplo 2.12

Siguiendo con el Ejemplo 2.1, ¿cómo podría ahora calcularse la función de distribución acumulada de la variable aleatoria número de semillas germinadas usando la función de densidad?

x	F(x)
0	F(0) = f(0) = P(0) = 1/8
1	F(1) = f(0) + f(1) = P(0) + P(1) = 1/8 + 3/8 = 4/8
2	F(2) = f(0) + f(1) + f(2) = P(0) + P(1) + P(2) = 1/8 + 3/8 + 3/8 = 7/8
3	F(3) = f(0) + f(1) + f(2) + f(3) =P(0) + P(1) + P(2) + P(3) = 1

Función de densidad de una variable aleatoria continua

Definición 2.14: Función de densidad de una v.a. continua

La **función de densidad** de una variable aleatoria continua es una función f(.) ≥ 0 tal que:

$$P(\ [\ x_1 \leq X \leq x_2]) = \int_{x1}^{x2} f(y)dy \,, \forall \ \ x_1, x_2 \in \Re \,.$$

Nota: Si X es una variable aleatoria continua, se dirá que la probabilidad de un valor cualquiera es cero, es decir, $P[X = x] = 0 \ \forall \ x$, ya que según la definición anterior, ésta correspondería al área asociada a un punto, la que geométricamente es nula.

De la definición de función de densidad de una variable aleatoria continua se deduce que la función de distribución acumulada F(.) de esta variable puede ser obtenida mediante la siguiente expresión:

$$F_X(x) = \int_{-\infty}^{x} f(y)d(y) \qquad \text{para cualquier número real x}$$

Es decir, para un x dado, $P([X \leq x]) = F(x)$ es el valor del área bajo la curva que representa a la función de densidad comprendida entre $-\infty$ y x.

Como consecuencia de lo expuesto, la probabilidad de que X tome un valor dentro de un intervalo determinado $[x_1, x_2]$ (ver Figura 2.4), puede obtenerse a partir de la función de distribución acumulada, mediante la siguiente relación:

$$P(A) = P([\ x_1 \leq X \leq x_2\ ]) = F(x_2) - F(x_1)$$

y dado que

$$F(x_2) = \int_{-\infty}^{x_2} f(y)dy \quad y \quad F(x_1) = \int_{-\infty}^{x_1} f(y)dy$$

Se tiene que

$$F(x_2) - F(x_1) = \int_{-\infty}^{x_2} f(y)dy - \int_{-\infty}^{x_1} f(y)dy = \int_{x_1}^{x_2} f(y)dy$$

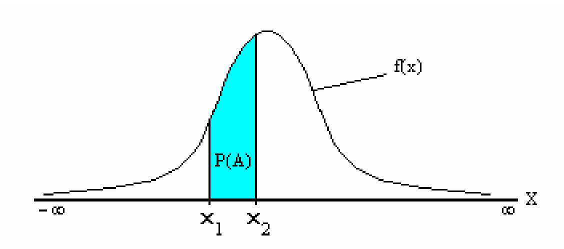


Figura 2.4: Gráfico de la función densidad f(x).

Notemos que $\int_{-\infty}^{\infty} f(y)dy = 1$, ya que, el intervalo (-∞,∞) contiene todos los posibles valores de la variable aleatoria, y la probabilidad de Ω es, por el axioma *i* de la definición de Kolmogorov, igual a 1.

Las propiedades de la función de distribución acumulada son las siguientes:

a) es no decreciente

b) es continua por derecha

c) para x → -∞ y x → +∞ los límites de F(x) son respectivamente 0 y 1.

Medidas Resumen de la Distribución de una Variable Aleatoria

Se ha visto que a cada variable aleatoria se le asocia una función de distribución. El estudio del comportamiento de una variable se puede realizar a partir de la descripción de su distribución. Para ello se utilizan ciertos valores o parámetros que la caracterizan. Por ejemplo, en la Figura 2.5, se muestran los gráficos de dos distribuciones cuyas diferencias radican en que, a pesar de poseer formas similares, una curva esta desplazada con respecto a la otra. Por ello es necesario contar con un parámetro que indique la posición de la curva sobre la recta real.

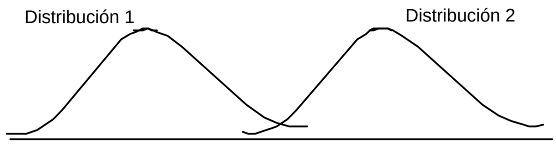


Figura 2.5: Gráfico de dos distribuciones con distintos parámetros de posición.

En otras distribuciones los valores más frecuentes, o con mayor densidad, pueden estar posicionados en el mismo intervalo y, sin embargo, las distribuciones ser distintas como muestra la Figura 2.6. Se observa en este caso, que bajo la distribución 2 los valores "alejados del centro" tienen mayor frecuencia que bajo la distribución 1.

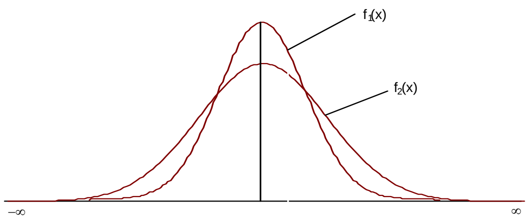


Figura 2.6: Gráfico de dos distribuciones con distinta variabilidad.

Como en los casos anteriores se pueden utilizar los gráficos de las funciones de densidad para la descripción del comportamiento de las variables aleatorias; en la práctica se usan medidas que resumen y cuantifican la información que se visualiza en los mismos. Las medidas de resumen más frecuentemente usadas son las llamadas medidas de posición y de dispersión.

Esperanza de una variable aleatoria

La esperanza matemática de una variable aleatoria es, desde un punto de vista intuitivo, un promedio de los valores asumidos por la variable, donde cada valor es "ponderado" por su probabilidad de ocurrencia.

Definición 2.15: Esperanza de una v.a. discreta

La **esperanza de una variable aleatoria discreta** X, con función de densidad f(.),es:

$$E(X) = \mu = \sum_{x_i \in C} x_i f(x_i)$$

donde C = { x_1, x_2, x_3, ... } es el conjunto de valores que puede tomar la variable aleatoria.

Definición 2.16: Esperanza de una v.a. continua

La **esperanza de una variable aleatoria continua** X, con función de densidad f(.), es:

$$E(X) = \mu = \int_{-\infty}^{\infty} x\, f(x)\; dx$$

Esta integral no siempre existe y en ese caso, se dirá que la variable no tiene esperanza.

Ejemplo 2.13

Siguiendo con el Ejemplo 2.1, si $x_1 = 0$, $x_2 = 1$, $x_3 = 2$ y $x_4 = 3$, representan los posibles valores de la variable aleatoria X = número de semillas sanas, la E(X) se calcula como sigue:

$$E(X) = x_1\, f(x_1) + x_2\, f(x_2) + x_3\, f(x_3) + x_4\, f(x_4)$$

$$= 0 \cdot \frac{1}{8} + 1 \cdot \frac{3}{8} + 2 \cdot \frac{3}{8} + 3 \cdot \frac{1}{8} = \frac{12}{8} = 1.5$$

Para una distribución dada, la esperanza es un valor constante que identifica al "centro de gravedad" de la función de densidad. Por ello, la esperanza puede ser vista como una medida que indica la posición de la distribución.

Volviendo a la Figura 2.5, se dice que la variable aleatoria cuya gráfica de la función de densidad es la curva 2 tiene una esperanza mayor que la asociada a la curva 1, lo cual se visualiza con el desplazamiento hacia la derecha del gráfico de la función de densidad.

Propiedades de la esperanza

Sean X_1 y X_2 v.a. con esperanzas definidas, *a* y *b* constantes; entonces se cumplen las siguientes propiedades:

a) $E(aX_1) = a\, E(X_1)$

b) $E(aX_1 + bX_2) = a\, E(X_1) + b\, E(X_2)$

Ejemplo 2.14

Siguiendo con el Ejemplo 2.13, ¿qué sucede con E(X) si se multiplica por 3 a X? Por la propiedad a) se tiene que E(3X) = 3E(X) = 3 (1.5). En efecto, por definición:

$$E(3X) = 3x_1\, f(x_1) + 3x_2\, f(x_2) + 3x_3 f(x_3) + 3x_4\, f(x_4)$$
$$= 3(0 \cdot \frac{1}{8} + 1 \cdot \frac{3}{8} + 2 \cdot \frac{3}{8} + 3 \cdot \frac{1}{8}) = 3 \cdot \frac{12}{8} = 3(1.5) = 4.5$$

Suponga que se está estudiando el número de cabritos por parición de dos razas de cabras y se asume que el número máximo de crías en una parición es de tres. Llamemos X_1 a la variable aleatoria número de cabritos por parición de la raza 1 y X_2 al número de cabritos por parición de la raza 2. Luego, $X_1 + X_2$ denota a la variable aleatoria "suma de los números de crías en una parición en ambas razas". El conjunto de resultados posibles esta representado en la siguiente tabla:

	$X_2 = 0$	$X_2 = 1$	$X_2 = 2$	$X_2 = 3$
$X_1 = 0$	(0,0) = 0	(0,1) = 1	(0,2) = 2	(0,3) = 3
$X_1 = 1$	(1,0) = 1	(1,1) = 2	(1,2) = 3	(1,3) = 4
$X_1 = 2$	(2,0) = 2	(2,1) = 3	(2,2) = 4	(2,3) = 5
$X_1 = 3$	(3,0) = 3	(3,1) = 4	(3,2) = 5	(3,3) = 6

En cada celda el primer elemento del par ordenado es un valor posible para X_1 y el segundo es un valor posible para X_2. Como X_1 es independiente de X_2 entonces $P(X_1 = x_1, X_2 = x_2) = P(X_1 = x_1)\, P(X_2 = x_2)$. Luego, sigue que:

$$\begin{aligned}
E(X_1 + X_2) = {} & 0\, P(0)\, P(0) + 1\, P(1)\, P(0) + 2\, P(2)\, P(0) + 3\, P(3)\, P(0) + \\
& + 1\, P(0)\, P(1) + 2\, P(1)\, P(1) + 3\, P(2)\, P(1) + 4\, P(3)\, P(1) + \\
& + 2\, P(0)\, P(2) + 3\, P(1)\, P(2) + 4\, P(2)\, P(2) + 5\, P(3)\, P(2) + \\
& + 3\, P(0)\, P(3) + 4\, P(1)\, P(3) + 5\, P(2)\, P(3) + 6\, P(3)\, P(3) \\
= {} & 0\,(1/16) + 1\,(1/16) + 2\,(1/16) + 3\,(1/16) + \\
& + 1\,(1/16) + 2\,(1/16) + 3\,(1/16) + 4\,(1/16) + \\
& + 2\,(1/16) + 3\,(1/16) + 4\,(1/16) + 5\,(1/16) + \\
& + 3\,(1/16) + 4\,(1/16) + 5\,(1/16) + 6\,(1/16) = \\
= {} & 0(1/16) + 1(2/16) + 2(3/16) + 3(4/16) + 4(3/16) + 5(2/16) + 6(1/16) \\
= {} & 3 = 1.5 + 1.5 \\
= {} & E(X_1) + E(X_2)
\end{aligned}$$

que es el resultado previsto por la propiedad *b* de la esperanza.

Varianza de una variable aleatoria

La esperanza de una variable aleatoria sólo proporciona información parcial acerca de su distribución, ya que explicita dónde está posicionada, pero, como se observa en la Figura 2.6, dos o más distribuciones pueden tener la misma medida de posición y sin embargo ser distintas. Existe la necesidad de una medida que indique cuan disímiles son los valores de la variable aleatoria. Antes de presentar la definición de varianza, considérese el siguiente ejemplo.

Ejemplo 2.15

Los dos conjuntos de datos, A_1, A_2, corresponden a todos los valores posibles e igualmente probables de dos variables aleatorias discretas X_1 y X_2 respectivamente. Ambas distribuciones tienen igual esperanza pero nótese la diferencia en su variación:

$$A_1 = \{8; 8; 9; 10; 11; 12; 12\} \quad \text{y} \quad A_2 = \{1; 2; 5; 10; 15; 18; 19\}$$

Si bien para las dos variables aleatorias, la esperanza es 10, para X_1 los posibles valores están más cerca de la esperanza que los de X_2. Por otro lado, X_1 tiene rango de variación igual a 4 y menor que el rango de X_2, que es igual a 18.

¿Es el rango suficiente para resumir la variabilidad de la variable aleatoria? Si se observan los conjuntos A_3 y A_4 de valores posibles de las v.a. X_3 y X_4, se tiene:

$$A_3 = \{8; 10; 10; 10; 10; 10; 12\} \quad \text{y} \quad A_4 = \{1; 10; 10; 10, 10; 10; 19\}$$

Se nota que también sus rangos son 4 y 18 y que la esperanza de ambas también es 10. Es decir, que la esperanza y el rango no hacen una caracterización satisfactoria de una variable aleatoria ya que se ve que distribuciones con igual rango y con igual esperanza, son realmente distintas. Las primeras distribuciones que se analizaron (distribución de X_1 y de X_2) presentan más dispersión de los datos que las últimas (distribución de X_3 y de X_4). Es deseable entonces, encontrar otra medida de dispersión distinta del rango.

Esta medida debería tiner un valor pequeño cuando la mayoría de las observaciones se encuentran cerca de la esperanza y un valor grande cuando estén muy alejadas.

Se define a continuación la varianza para una variable aleatoria discreta y para una variable aleatoria continua, tal cual se hizo para el caso de la esperanza.

Definición 2.17: Varianza de una v.a. discreta

La **varianza de una variable aleatoria discreta X** se define como:

$$Var(X) = \sigma^2 = \sum_{x_i \in C} (x_i - \mu)^2 f(x_i)$$

donde μ = E(X), f(.) la función de densidad y C = $\{x_1, x_2, ...\}$ el conjunto de valores posibles.

Definición 2.18: Varianza de una v.a. continua

La **varianza de una variable aleatoria continua** X, denotada por V(X) ó σ^2, es:

$$V(X) = \sigma^2 = \int_{-\infty}^{\infty} \left(x - \mu\right)^2 f(x)d(x),$$

donde μ = E(X)) y f(.) la función de densidad.

Cabe señalar que esta integral no siempre existe, en este caso se dirá que la v.a. no tiene varianza.

Propiedades de la Varianza

Sean X_1 y X_2 variables aleatorias, *a* y *b* constantes. Entonces, se cumple que:

a) Si $X_1 = a$, entonces V(X) = 0

b) $V(aX) = a^2V(X)$

c) $Var(X_1 + a) = Var(X_1)$

d) $Var(aX_1 \pm bX_2) = a^2Var(X_1) + b^2Var(X_2) \pm ab\, 2Cov(X_1, X_2)$

¿Qué representa el valor de la varianza de una variable aleatoria? La varianza es un "promedio ponderado" de los cuadrados de los desvíos respecto de la esperanza.

Volviendo a la Figura 2.6, se dice ahora que la variable aleatoria cuyo gráfico de la función de densidad es la curva 2 tiene una varianza mayor que la asociada a la curva 1.

Debido a que la varianza se expresa en función de desvíos al cuadrado, el valor de la varianza no está en la escala original en que están expresadas las observaciones de la variable aleatoria. Es decir que si X se mide en cm, entonces la V(X) se expresa en cm^2 .

Un parámetro de dispersión en la escala original se obtiene tomando la raíz cuadrada de la varianza y se conoce como **desvío estándar** o **desviación estándar** o **desviación**

típica.

$$\sigma = \sqrt{\sigma^2}$$

Otra medida de dispersión usada para referirse a la distribución de una variable aleatoria es el coeficiente de variación. Este tiene la ventaja de independizarse de la escala original de los datos expresando la variabilidad en forma relativa a la magnitud de la esperanza.

Definición 2.19: Coeficiente de variación.

El **coeficiente de variación** de una variable aleatoria con esperanza μ y desvío estándar σ es:

$$CV = \frac{\sigma}{\mu} \cdot 100$$

El coeficiente de variación expresa la desviación estándar como porcentaje respecto de la esperanza. Es útil para comparar la variabilidad de dos o más variables aleatorias, expresadas en diferentes unidades de medida.

De este modo, con los valores de esperanza y varianza se puede resumir aspectos relevantes del gráfico de la función densidad, por ende de la función de distribución de la variable aleatoria.

Sin embargo, dos distribuciones pueden tener igual esperanza y varianza y ser aún diferentes. Otras medidas de resumen, tales como la asimetría y la kurtosis son usadas para describir la forma de las distribuciones.

Para una variable aleatoria continua se dice que una distribución es simétrica, respecto a un eje que pasa por la esperanza, si el área bajo la curva de la función de densidad "a la derecha" es una imagen especular del área a la izquierda, en caso contrario diremos que es asimétrica. La asimetría podrá ser derecha o izquierda como se ilustra en la Figura 2.7.

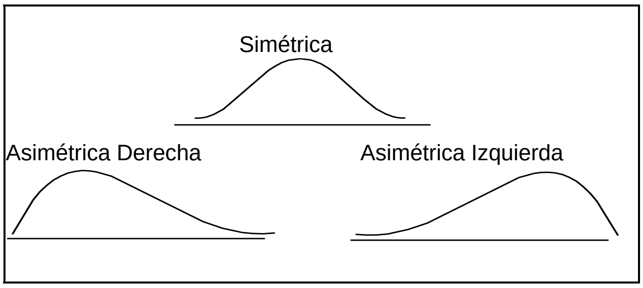


Figura 2.7: Gráficas de funciones densidad con asimetría de naturaleza distinta.

Cuantiles de una variable aleatoria

Un concepto muy usado para el estudio de variables aleatorias y sus distribuciones es el de cuantil. Este se encuentra estrechamente ligado a la función de distribución acumulada. A continuación se presenta la definición de cuantil para variables aleatorias continuas (Conover, 1980).

Definición 2.20: Cuantil

Si X es una variable aleatoria continua, el **cuantil** x_p se define como el valor x tal que:

$$P[X \leq x_p] = p$$

Por ejemplo, si $P[X \leq x] = 0.10$ entonces x es el cuantil 0.10 de la variable X.

Ejemplo 2.16

Si se desea conocer el tanto por ciento de espigas de trigo que tienen una longitud menor o igual a 12 cm, es equivalente a querer conocer a qué cuantil corresponde el valor 12 de la variable "longitud de espigas de trigo".
Es común encontrar el uso de la noción de cuantil expresada como porcentaje, así el cuantil $x_{0.10}$ pasa a denominarse el percentil 10, el cuantil $x_{0.75}$ el percentil 75, etc.

Los cuantiles o los percentiles son generalmente usados para fijar límites de tolerancia para los valores de algunas variables. En medicina, los límites normales de talla o de peso para un niño de dos años, no son más que los cuantiles 0.05 y 0.95 de la talla o del peso en la población de niños normales de esa edad.

Ejercicios

Ejercicio 2.1

El espacio muestral para un experimento aleatorio en el cual se estudia la parición simultánea de dos conejas, cada una de las cuales puede tener como máximo 6 crías y siempre tiene al menos una cría, es el siguiente:

$$\Omega = \{ (x,y) / x = 1,2,....,6; \wedge y = 1,2,...,6\}$$

a) Describir este espacio que está constituido por los 36 elementos o puntos muestrales, cada uno representado por el par (x,y), donde x = número de crías de la coneja 1 e y = número de crías de la coneja 2.

b) ¿El espacio Ω es finito o infinito?

c) ¿Se puede decir que el total de crías es una variable aleatoria? ¿De qué tipo?

Nota: Algunas veces, el conjunto de resultados posibles de un experimento no es tan fácil de definir. Por ejemplo, esto ocurre al seleccionar al azar un habitante de la ciudad de Córdoba y medir su altura en metros. En este caso, ¿cuáles son los resultados posibles del experimento?, ¿Son los números reales entre 0 y ......? Suponiendo que no existe una altura máxima, tal vez sea razonable elegir $\Omega=(0; \infty)$, no obstante se sabe que este conjunto contiene resultados imposibles, como 1.000.000 de metros. Otros conjuntos candidatos para Ω podrían ser los siguientes: (0 ; 3), (1/10 ; 3). Estos dos intervalos contienen aparentemente todos los resultados posibles.

Ejercicio 2.2

Describir el espacio muestral asociado al experimento: "registrar la parición de una conejera hasta que un conejo con malformación física nazca". Si en una parición no sucede lo esperado se representa la misma con la letra N (normal) y si la malformación ocurre se usa la letra M (malformado).

a) ¿Contiene este espacio un número finito o infinito numerable de elementos?

b) ¿Qué variable aleatoria se puede definir sobre el mismo? Caracterizarla.

Ejercicio 2.3

Con referencia al espacio muestral del Ejercicio 2.1, describir el evento A: "que al menos una coneja sea mellicera" y el evento B: "el número total de crías no supera 5".

Ejercicio 2.4

Un investigador que estudia métodos de aplicación de nutrientes foliares necesita indagar sobre la localización de las gotas de nutrientes sobre la hoja cuando el fertilizante se aplica con un método de aspersión. Para ubicar las gotas supóngase que la hoja es un plano y que se puede definir precisamente las coordenadas del punto sobre el que cae la gota. Describir el espacio muestral de este experimento.

Ejercicio 2.5

Supóngase que se conduce una investigación para determinar la distancia que podría recorrer un tractor sobre un camino, con 5 litros de gas-oil en ciertas condiciones de avance. La distancia es una variable aleatoria que puede ser medida con el grado de precisión deseada.

a) Describir el espacio muestral de este experimento.

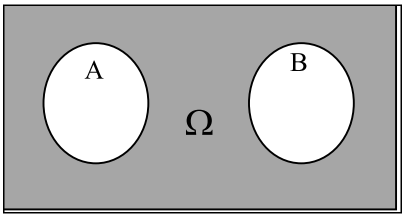
b) Identificar el tipo de variable en estudio.

Ejercicio 2.6

Construir el espacio muestral asociado al tiempo (t) de vida útil de un componente electrónico (medida en años), y señalar el subconjunto que representa al siguiente evento: "que el componente falle antes del final del sexto año".

Ejercicio 2.7

El siguiente rectángulo representa un espacio muestral y los eventos son representados por regiones (círculos) dentro del rectángulo.



a) ¿Son estos eventos mutuamente excluyentes?

b) Graficar una situación donde ambos eventos pueden ocurrir simultáneamente.

c) Bajo la situación dada en b), ¿cuál es la probabilidad de que suceda A ó B?

Ejercicio 2.8

Un productor tambero desea aumentar el número de vacas lecheras de su tambo en un período de dos años. Para esto necesita conocer: a) ¿cuál es la probabilidad de tener al menos una cría hembra por vaca en las dos pariciones considerando una producción de 1 ternero por vaca por año y que la proporción de sexos es 1:1? b) ¿Cuál es la probabilidad de que teniendo 20 vacas no nazca ninguna hembra?

Ejercicio 2.9

Para cada una de las siguientes situaciones, explicar por qué ellas no son formas permisibles de asignar probabilidades a los 4 eventos posibles y mutuamente excluyentes A, B, C y D de un experimento aleatorio.

Situación A: P(A) = 0.12   P(B) = 0.63   P(C) = 0.45   P(D) = -0.20

Situación B: $P(A) = \frac{9}{120}$ $P(B) = \frac{45}{120}$ $P(C) = \frac{27}{120}$ $P(D) = \frac{46}{120}$

Ejercicio 2.10

Con el siguiente espacio muestral Ω = {HM, HH, MH, MM}, construir:

a) Dos eventos que no sean excluyentes ni independientes.

b) Dos eventos mutuamente excluyentes.

c) Dos eventos independientes.

Ejercicio 2.11

En un experimento para control de calidad de tractores, se le da arranque a las unidades en 4 oportunidades. En cada caso pueden arrancar (éxito) o no (fracaso).

a) Construir el espacio muestral.

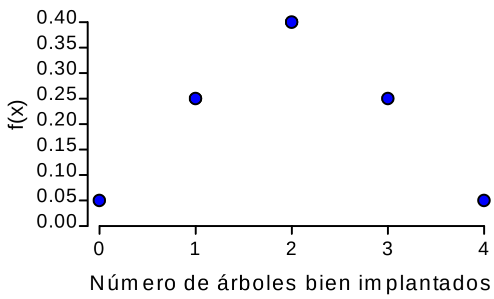
b) Asumiendo que todos los eventos elementales poseen la misma probabilidad, ¿cuál sería ese valor?

c) Listar los posibles valores de la variable aleatoria X definida como el número total de arranques exitosos.

d) ¿Cuál es la $P(X = 3)$? ¿Cuál es la $P(X \leq 2)$?

Ejercicio 2.12

La función de densidad de la variable aleatoria definida como el "*número de árboles bien implantados*", tiene la siguiente forma:



a) Construir la función de distribución acumulada correspondiente.

b) ¿Qué significa F(2)?

c) Calcular la media y la varianza poblacionales.

Ejercicio 2.13

Se conoce que el cuantil 0.10 de la distribución de la variable X = *longitud de raíces* de plántulas de tomate al momento del transplante es 3 cm, y se sabe que sólo las plántulas con raíces mayores de 3 cm tienen probabilidad de sobrevivir al transplante:

¿Cuántas plántulas se deberían adquirir para lograr un lote de 2000 plántulas implantadas?

Ejercicio 2.14

Dibujar, a mano alzada, densidades de variables aleatorias continuas, que sean:

a) Una simétrica y una asimétrica.

b) Con alta densidad de valores concentrados en torno de la esperanza.

c) Dos distribuciones, una con mayor varianza que la otra.

d) Una distribución con concentración de valores en dos puntos.

3

Modelos Estadísticos: Distribución Normal y Otras Distribuciones

Introducción

El concepto de variable aleatoria está íntimamente ligado al de función de densidad y función de distribución. Por lo general la forma o expresión matemática de la función que describe a la variable aleatoria no se conoce, por lo que los técnicos e investigadores suelen proceder a recolectar datos mediante estudios observacionales o experimentales, y a partir de ellos buscar cuál es la *función que mejor describe* la o las variables aleatorias en estudio.

No cualquier función matemática es útil para caracterizar una variable aleatoria, por el contrario, las funciones de densidad y de distribución acumulada deben reunir una serie de propiedades para que sea posible asignar probabilidades a los eventos de interés a partir de las mismas. Desde el punto de vista teórico se han estudiado con suficiente detalle un conjunto de funciones matemáticas que verifican las propiedades de las funciones de distribución acumulada y de las funciones de densidad tanto para variables discretas como para continuas. Luego, el técnico o investigador que no conoce la función exacta que caracteriza a la variable aleatoria que está estudiando puede, por *conocimiento empírico*, proponer alguna de las funciones, del conjunto de funciones antes indicado, para describir el comportamiento de su variable. De la habilidad para escoger una distribución adecuada, depende la calidad de los modelos y las predicciones que se construyan.

Si la selección de la función se realiza a partir de la distribución empírica de la variable (distribución de los valores muestrales), cuanto mayor sea el conjunto de datos recolectados, se podrá realizar una mejor identificación de la función.

Suele ocurrir que, aunque los datos hayan sido bien tomados, el conjunto seleccionado no sea bien descripto por alguna de las funciones conocidas por lo que las conclusiones del estudio siempre dependerán del grado de aproximación logrado.

Un **modelo** se define como una representación simplificada de la realidad. En el estudio de una variable aleatoria se utiliza el término modelo para hacer referencia a la función de distribución seleccionada aunque ésta no sea la que caracteriza exactamente el comportamiento de la variable aleatoria.

Al proceso de selección del modelo distribucional se lo conoce como **modelación**. Sin embargo, el concepto de modelo también tiene un significado más amplio tanto en la misma estadística como en matemática y otras ciencias. La modelación es la base de la inferencia estadística, es decir, el procedimiento inductivo mediante el cual, a partir de las observaciones realizadas, se describen las características de la distribución bajo estudio. En general, varios modelos con diferentes niveles de complejidad pueden ser propuestos para el mismo problema y la adopción de uno u otro depende no sólo del grado de conocimiento que se tiene sobre la característica que se está investigando sino también de los **objetivos** que se persiguen.

La modelación también es usada para estudiar cuáles son y qué magnitud relativa presentan las distintas fuentes de variación de una variable aleatoria. Es decir, qué factor hace que la variable en estudio cambie o varíe, lo cual es explicitado a través de un modelo matemático.

En este Capítulo estudiaremos la función de densidad normal o modelo de Gauss, que permite aproximar el comportamiento estadístico de muchas variables continuas e incluso de algunas variables discretas. La distribución normal es un modelo de probabilidad y una vez adoptado el modelo es posible responder a las siguientes preguntas:

-*¿Cuál es la probabilidad de que la variable en estudio tome valores menores a un valor determinado?*

Por ejemplo, si la variable es el rendimiento de un cultivar, el responder a esta pregunta podría indicar la posibilidad de obtener rendimientos que no justifiquen el costo de producción.

-*¿Cuál es la probabilidad de que la variable en estudio tome valores mayores a un valor determinado?*

Si la variable aleatoria en estudio es la cantidad de semillas de maleza en el suelo antes de la siembra, el responder a esta pregunta podría indicar si se necesitará o no aplicar herbicida (este podría ser el caso de modelación de una variable aleatoria discreta como si se tratara de una continua).

-*¿Cuál es la probabilidad de que la variable en estudio tome valores entre 2 valores determinados?*

Se podrían mencionar un gran número de ejemplos, tanto de la práctica como de la investigación agronómica en los que responder a esta pregunta podría ser de interés.

Suponer una distribución determinada para una variable aleatoria servirá además, como se verá más adelante, para realizar pruebas que permitan rechazar o mantener hipótesis postuladas en el marco de la investigación. La base para el establecimiento de esas conclusiones serán los modelos probabilísticos, en relación a los cuales se especifican las cuestiones de interés. A continuación se presenta y discute en detalle la distribución normal. Posteriormente se presentan otras distribuciones para variables continuas y discretas.

Distribución Normal

La Función de Densidad Normal

Esta función, también conocida como "campana de Gauss", desempeña un papel central en la teoría y la práctica de la estadística. Muchos fenómenos agronómicos, biológicos, químicos, físicos, antropológicos, etc., son estudiados a partir de datos distribuidos de manera normal. Variables continuas, tales como peso, longitud, altura, temperatura, absorbancia óptica, resistencia a la tracción, etc. presentan gráficas de distribuciones de frecuencias que se pueden aproximar muy bien por esta función de densidad.

Definición 3.1: Variable aleatoria normal

Una **variable aleatoria** X se define como normalmente distribuida si su función de densidad está dada por:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

donde: los parámetros μ y σ satisfacen $-\infty \leq \mu \leq \infty$ y $\sigma>0$

e = base de los logaritmos naturales (aprox: 2.7182818), π = constante matemática aproximada por 3.14159 y x $\in$ $(-\infty, \infty)$.

La representación gráfica de la función de densidad normal es una curva simétrica que tiene forma de campana (Figura 3.1). La localización del centro de la campana está dado por el parámetro μ (la esperanza) y la mayor o menor amplitud de la campana viene dada por σ^2 (la varianza).

Nota: Como la función es simétrica respecto de μ, ésta divide a la gráfica en partes iguales. Está definida para todo $\Re$ y para valores en la abscisa que tienden a infinito y menos infinito, se aproxima al eje horizontal sin tocarlo (curva asintótica). Como toda función de densidad, el área comprendida entre el eje de las abscisas y la curva es igual a la unidad.

Si se fijan dos puntos cualesquiera, por ejemplo x_1 y x_2, sobre el eje que representa los valores de la variable (abscisas), la porción del área por debajo de la curva que queda comprendida entre esos dos puntos corresponde a la probabilidad de que la variable aleatoria *se realice* entre x_1 y x_2.

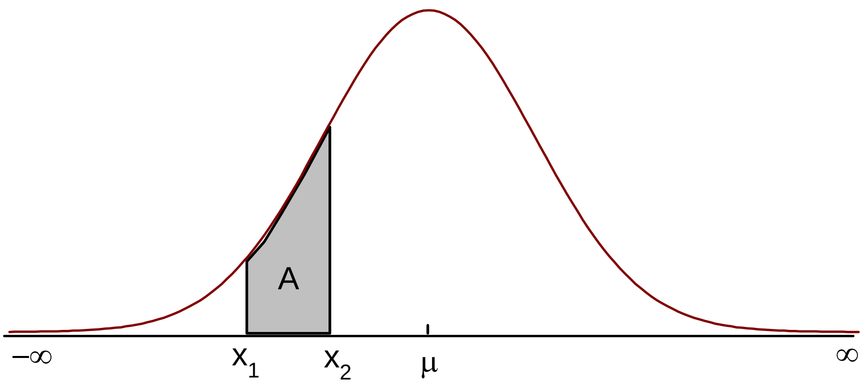


Figura 3.1: En las abscisas, posibles valores que puede tomar la variable aleatoria X y en las ordenadas valores de la función de densidad normal; A representa la probabilidad asociada con valores de X comprendidos entre x_1 y x_2

Si se llama A a esta área, se puede representar simbólicamente lo expuesto anteriormente como:

$$A = P\ (x_1 \leq X \leq x_2)$$

Existen infinitas distribuciones normales. Cada una de ellas queda especificada por los parámetros μ y σ^2. Es por ello que cuando se quiere indicar que una variable X tiene distribución normal caracterizada por μ (esperanza) y σ^2 (varianza) se escribe:

$$X \sim N\ (\mu,\sigma^2)$$

La Figura 3.2 presenta dos densidades normales con distinta varianza.

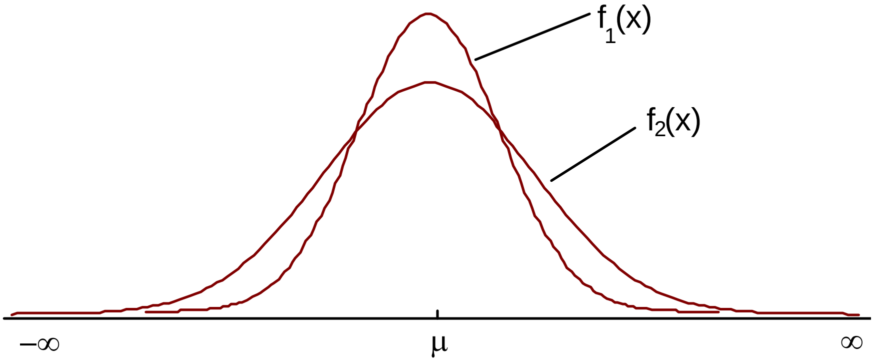


Figura 3.2: Dos densidades normales con igual media μ pero diferente varianza σ^2. Observar que $\mu_1 = \mu_2$ y $\sigma_1{}^2 < \sigma_2{}^2$

La mayor densidad se encuentra para valores x cercanos a μ y los puntos de inflexión están en μ - σ y μ + σ. La Figura 3.3 presenta algunos ejemplos de densidades normales variando la esperanza.

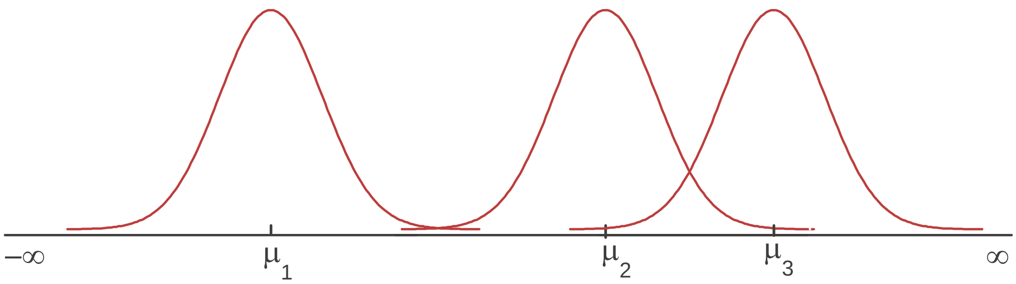


Figura 3.3: Tres densidades normales con igual desviación estándar pero diferentes medias

Por las propiedades de las funciones de densidad de variables continuas, si se quiere conocer **la probabilidad de que una variable distribuida normalmente "se realice" entre x_1 y x_2** (es decir, conocer la magnitud del área A citada anteriormente) se deberá integrar entre x_1 y x_2 la función de densidad normal, de la siguiente manera:

$$P(x_1 \leq X \leq x_2) = \int_{x_1}^{x_2} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx$$

con μ y σ conocidas. Esta integral no tiene una expresión analítica y por lo tanto se debe resolver numéricamente.

Estandarización

Para entender este concepto se plantea la siguiente situación: supóngase que la longitud de las alas de la mosca de los frutos tiene función de densidad normal y que la longitud de las alas de gallinas también. Esto no quiere decir que la función de densidad de la variable longitud de las alas de las moscas de los frutos sea igual a la de la longitud de las alas de gallinas ya que, obviamente, los parámetros de ambas funciones de densidad son distintos. Es de esperar que el promedio de longitudes de las alas de moscas sea menor al promedio de longitudes de alas de gallinas.

Luego, a pesar de que muchas variables puedan presentar funciones de densidad aproximadamente normales, cada problema se asocia con una función de densidad normal diferente por lo que, en cada caso, si se desea conocer la probabilidad de que la variable tome ciertos valores, se debería integrar una función de densidad diferente (entiéndase una función normal con diferentes parámetros).

Este trabajo se simplifica usando una **transformación** que hace que variables aleatorias con funciones de densidad normal diferentes, se distribuyan de la misma manera bajo la transformación, facilitando así los cálculos de probabilidades con cualquier combinación de parámetros μ y σ^2.

Definición 3.2: Estandarización

Se llamará **estandarización** a la siguiente transformación:

$$Z = \frac{X - \mu}{\sqrt{\sigma^2}}$$

donde : **Z:** es la variable aleatoria obtenida de la transformación

X: la variable aleatoria original

μ y σ^2 son respectivamente, la esperanza y la varianza de la distribución de X.

Definición 3.3: Función de densidad normal estándar

Se llamará **función de densidad normal estándar** y se simbolizará como N (0,1) a:

$$f(z) = \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-\frac{1}{2}Z^2}$$

Si **X** se distribuye normal con media μ y varianza σ^2, luego **la variable Z** (la estandarización de X), se distribuye normal con media 0 y varianza 1, esto es:

$$X \sim N(\mu,\sigma^2) \Longrightarrow Z = \frac{X-\mu}{\sqrt{\sigma^2}} \sim N(0,1)$$

Se ha reducido el problema de tener muchas distribuciones, a tener una sola. Pero para hallar la probabilidad de que X tome un valor entre dos valores determinados se deberá aún integrar la función de densidad N (0,1).

Ejemplo 3.1

Si $X \sim N(\mu,\sigma^2)$ con $\mu = 10$ y $\sigma^2 = 4$ y se desea conocer la $P[\,8 \le X \le 9\,]$ se procede de la siguiente manera:

a) Se estandariza de modo que queda: $z_1 = \frac{8-10}{2} = -1$ y $z_2 = \frac{9-10}{2} = -0.5$

b) Luego: $A = P[\,8 \le X \le 9\,] = P[\,-1 \le Z \le -0.5\,] = B$, ilustrado en la siguiente figura:

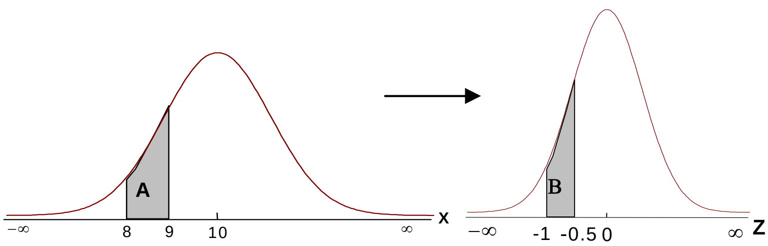


Figura 3.4: La transformación de estandarización

c) Calcular $B = P[\,-1 \le Z \le -0.5\,]$ como se explica a continuación.

Para hallar la solución a este problema, es decir, para encontrar el valor del área sombreada en el gráfico anterior, deberíamos resolver la siguiente integral:

$$\int_{z_1}^{z_2} \frac{1}{\sqrt{2\pi}}\, e^{-\frac{1}{2}z^2}\, dz = \int_{-\infty}^{z_2} \frac{1}{\sqrt{2\pi}}\, e^{-\frac{1}{2}z^2}\, dz - \int_{-\infty}^{z_1} \frac{1}{\sqrt{2\pi}}\, e^{-\frac{1}{2}z^2}\, dz$$

Afortunadamente, las integrales de la forma: $\int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} dz$ están calculadas, para un conjunto grande y usualmente suficiente de valores de z entre -3.5 y +3.5 que se pueden encontrar en la tabla de cuantiles de la Distribución Normal Estándar (Apéndice).

En la actualidad, es muy simple generar estas tablas a partir de funciones estadísticas de las planillas de cálculo.

Función de Distribución Acumulada Normal

La función de distribución acumulada de una variable aleatoria normal, evaluada en el punto *x* (F(x)), describe la probabilidad de que la variable tome valores menores o iguales a *x*. Esto es, en el caso normal:

$$P(-\infty \leq X \leq x) = F(x)$$

Luego, utilizando propiedades de integrales, la **P[$x_1 \leq X \leq x_2$],** puede ser resuelta como: **F(x_2) - F(x_1)**, como se ilustra en la Figura 3.5.

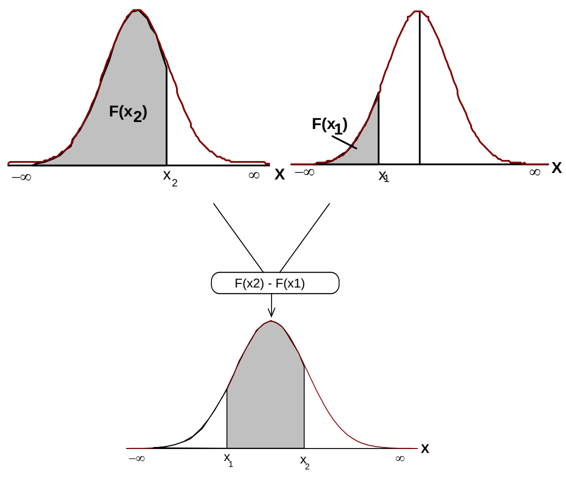


Figura 3.5: Representación gráfica de la $P[x_1 \leq X \leq x_2]$ vista como diferencia entre $F(x_2)$ y $F(x_1)$.

Para conocer cuánto vale $F(x_2)$ y $F(x_1)$, se procede de la siguiente manera:

1. Estandarizar el valor x_1 y el valor x_2, haciendo:

$$z_1 = \frac{x_1 - \mu}{\sqrt{\sigma^2}} \quad \text{y} \quad z_2 = \frac{x_2 - \mu}{\sqrt{\sigma^2}}$$

2. Como se vio, P [$x_1 \le X \le x_2$] = P [$z_1 \le Z \le z_2$], y en términos de la función de distribución es equivalente a $F(z_2)$ - $F(z_1)$. Luego, usando la tabla correspondiente a la función de distribución normal acumulada se hallan $F(z_1)$ y $F(z_2)$

Resumiendo, $P(x_1 \le X \le x_2) = P(z_1 \le Z \le z_2) = F(z_2) - F(z_1)$

Ya que $P(x_1 \le X \le x_2) = \int_{x_1}^{x_2} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dz$

con μ y σ conocidos, e y π constantes.

Si X ~ N $(\mu,\sigma^2) \Rightarrow z = \frac{x-\mu}{\sigma}$ ~ N (0,1); con lo cual:

$$P(z_1 \le Z \le z_2) = \int_{z_1}^{z_2} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} dz =$$

$$= \int_{-\infty}^{z_2} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} dz - \int_{-\infty}^{z_1} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} dz$$

Lo que es igual a: $F(z_2) - F(z_1)$

Ejemplo 3.2

Sea X ~ N (μ = 10, σ^2 = 4). Calcular P [8 ≤ X ≤ 9]

1) $z_1 = \frac{8-10}{2} = -1$ y $z_2 = \frac{9-10}{2} = -0.5$

2) P[8 ≤ X ≤ 9] = P [-1 ≤ Z ≤ -0.5], en la Tabla Normal entrando por la columna que presenta los valores de z y buscando el valor -0.5, en la columna vecina se leerá el valor correspondiente a $F(z_2)$ que para el ejemplo es F(-0.5) = 0.3085. De la misma manera se halla la $F(z_1)$; en este ejemplo F(-1) = 0.1587. Calculando F(-0.5) - F(-1) = 0.3085 - 0.1587 = 0.1498

Se concluye que para la variable aleatoria X que se distribuye normalmente con μ= 10 y σ^2 = 4, la probabilidad de que X se realice entre 8 y 9 es de 0.1498. Es decir que el área por debajo la curva de la función de densidad normal, caracterizada por los mencionados parámetros, y que se extiende por encima del segmento delimitado por los valores 8 y 9, corresponde al 14.98 % del área total bajo la curva.

Ejemplo 3.3

Supóngase que la variable en estudio tiene distribución normal, con $\mu = 5$ y $\sigma^2 = 4$ y se quiere conocer la probabilidad de que la variable tome valores mayores a 7.78.

Conociendo que P [X ≥ 7.78] puede reescribirse como 1-P[X ≤ 7.78], equivalente a 1 - F (7.78), entonces:

1) Se calcula el valor z para x = 7.78

$$z = \frac{x - \mu}{\sqrt{\sigma^2}} = \frac{7.78 - 5}{\sqrt{4}} = 1.39$$

2) Se busca en la Tabla Normal la probabilidad para z = 1.40 (por aproximación a 1.39), es decir F(1.40). Para z = 1.40 el valor de probabilidad presentado en la tabla es 0.9192. Luego, P(X > 7.78)=P(Z > 1.4) = 1 - 0.9192 = 0.0808, cuya representación gráfica es:

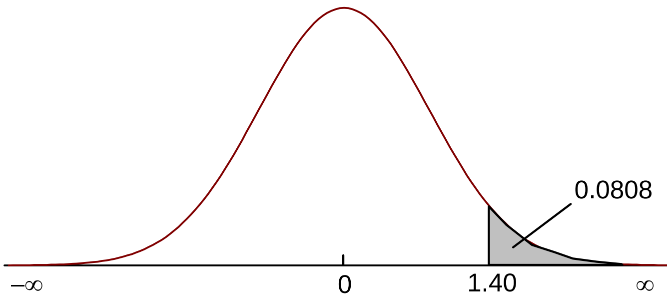


Figura 3.6: Representación del área bajo la curva normal por encima del valor 1.4

La variable Z puede ser vista como una desviación de X en torno a la media medida en unidades de desviación estándar. Es decir P [-1 < Z < 1] debe entenderse como la probabilidad de que X tome valores que se alejan de la media en menos o más una desviación estándar, es decir, P [μ - 1σ < X < μ + 1σ].

En una distribución normal teórica, esta probabilidad es igual a 0.6827, lo que equivale a decir que en la distribución normal el 68.27% de las observaciones están comprendidas entre la esperanza menos un desvío estándar y la esperanza más un desvío estándar:

$[\mu \pm 1\ \sigma]$ incluye al 68.27% de las observaciones

De igual manera se deduce que:

$[\mu \pm 2\ \sigma]$ incluye al 95.45% de las observaciones

$[\mu \pm 3\ \sigma]$ incluye al 99.74% de las observaciones

Existen pruebas formales para verificar el supuesto de normalidad que se pueden aplicar a una distribución empírica. Estas técnicas no serán desarrolladas en el marco de esta obra.

Otras distribuciones

Funciones de densidad de variables aleatorias discretas

En la presentación de cada función se seguirá el siguiente estilo y secuencia: a) situaciones en las que se puede seleccionar la función como modelo, b) definición de la función, c) propiedades, *i.e.* los parámetros que la caracterizan y d) ejemplos.

Se presentará una secuencia en complejidad e integración de conceptos crecientes, iniciando la misma, con funciones muy sencillas, pero necesarias para comprender las siguientes.

Distribución Uniforme Discreta

Se denotará a los posibles valores que pueda tomar una variable aleatoria discreta como x_1, x_2,..., x_k. En aquellos casos en que la variable aleatoria en estudio puede tomar sólo k valores con igual probabilidad cada uno de ellos, se dice que la variable aleatoria tiene distribución uniforme discreta. Con ello se quiere decir que la función de densidad de la variable aleatoria considerada es uniforme (constante).

Definición 3.4: Distribución Uniforme Discreta.

Una variable aleatoria X tiene distribución **Uniforme Discreta** si y sólo si su función de densidad es:

$$f(x) = \begin{cases} \frac{1}{k}\ para \quad x = x_1,\ x_2,\ ...,\ x_k \\ 0 \quad en\ caso\ contrario \end{cases}$$

donde $x_i \neq x_j$ para $i \neq j$.

Las variables aleatorias uniformes discretas se indican con la siguiente notación:

$$\mathbf{X} \sim Ud(x_1, x_k)$$

La esperanza E(**X**) y la varianza V(**X**) cuando **X** tiene distribución uniforme discreta se calcula como:

$$\mu = E(X) = \Sigma_{i=1}^{k} x_i f(x_i) = \Sigma_{i=1}^{k} x_i \frac{1}{k} = \frac{1}{k} \Sigma_{i=1}^{k} x_i = \frac{1}{k}(x_1 + x_2 + \ldots + x_k)$$

$$\sigma^2 = V(X) = \Sigma_{i=1}^{k} (x_i - \mu)^2 f(x_i) = \Sigma_{i=1}^{k} (x_i - \mu)^2 \frac{1}{k} = \frac{1}{k} \Sigma_{i=1}^{k} (x_i - \mu)^2 =$$

$$= \frac{1}{k}((x_1 - \mu)^2 + (x_2 - \mu)^2 + \ldots + (x_k - \mu)^2)$$

Ejemplo 3.4

Esta distribución aparece asociada a muchos juegos de azar, en los que los resultados tienen idéntica chance de ocurrir. Este es el caso de la ruleta, la quiniela, etc.

Esta distribución se usa, en el contexto del diseño de experimentos, para la asignación, con idéntica probabilidad, de las unidades experimentales a los tratamientos que se quieren comparar. En el contexto del muestreo, para seleccionar, con idéntica probabilidad, las unidades muestrales que conforman una muestra.

Distribución Bernoulli

En ciertos experimentos suele ocurrir que existen sólo dos resultados posibles: éxito o fracaso, presencia o ausencia, sí o no, etc.

En estos casos, se puede asociar a cada uno de los resultados posibles el número 0 o el número 1, según convenga. Por ejemplo, si el resultado de interés es el "éxito", se podría tomar $x = 1$ y si es "fracaso" hacer $x = 0$. Si el resultado de interés fuera el "fracaso", luego se debería asignar al revés.

Por otro lado, como el resultado del experimento es aleatorio, será natural pensar que cada uno de los resultados posibles tendrá cierta probabilidad de ocurrencia. En ciertas circunstancias ambos resultados pueden tener la misma probabilidad, pero obviamente no siempre es así. Si se llama θ a la probabilidad de uno de los dos resultados, luego la probabilidad del otro será 1-θ.

Definición 3.5: Distribución Bernoulli

Una variable aleatoria X tiene **distribución Bernoulli** si y sólo si su función de densidad es:

$$f(x;\theta) = \begin{cases} \theta^{x} \quad (1\text{-}\theta)^{1\text{-}x} \text{ si x} = 0\ ,1 \\ 0 \quad \textit{en caso contrario} \end{cases}$$

donde $0 \leq \theta \leq 1$.

Se denota a las variables Bernoulli con parámetro θ como **X** ~ *Ber*(θ).

Nota: cuando se escribe f(x;θ) se denota que x es el argumento de la función y que lo que sigue a continuación del punto y coma es una constante previamente especificada, necesaria para poder hacer cálculos con la función. Así, si θ = 0.3, luego la función de densidad Bernoulli será $f(x\ ;\ 0.3)=0.3^{x}\,(1\text{-}0.3)^{1\text{-}x}$.

La E(**X**) y la V(**X)** cuando **X** tiene distribución Bernoulli se calculan como:

$$\mu = E(X) = \Sigma_{x=0;1} x\, f(x) = \Sigma_{x=0;1} x\,(\theta^{x}(1-\theta)^{1-x}) = 0\,(\theta^{0}(1-\theta)^{1-0}) + 1\,(\theta^{1}(1-\theta)^{1-1}) =$$
$$= 1(\theta^{1}(1-\theta)^{1-1}) = \theta$$

$$\sigma^{2} = V(X) = \Sigma_{x=0;1}(x-\mu)^{2} f(x) = \Sigma_{x=0;1}(x-\mu)^{2}(\theta^{x}(1-\theta)^{1-x}$$

Como μ = θ, reemplazando, sigue que:

$$= (0-\theta)^{2}(\theta^{0}(1-\theta)^{1-0} + (1-\theta)^{2}(\theta^{1}(1-\theta)^{1-1}$$

Desarrollando los cuadrados y los exponentes, sigue:

$$= \theta^{2}(1-\theta) + (1^{2} - 2\theta + \theta^{2})\theta = \theta^{2} - \theta^{3} + \theta - 2\theta^{2} + \theta^{3} = \theta - \theta^{2} = \theta(1-\theta)$$

Nota: Obsérvese que θ caracteriza completamente a la función de densidad Bernoulli, es decir que tanto su esperanza como su varianza son expresiones que sólo dependen de θ. En este sentido se dice que θ es el "único parámetro" de esta función de distribución discreta.

Ejemplo 3.5

Presencia o ausencia de enfermedades en una planta, clasificación de semillas en anormales y normales, son ejemplos de variables aleatorias que se pueden modelar con una distribución Bernoulli.

Ensayos o experimentos en los que interesa el estudio de una o más variables aleatorias Bernoulli, son llamados *Ensayos o Experimentos Bernoulli*.

Distribución Binomial

Esta distribución tiene origen cuando ocurren las siguientes tres condiciones en forma simultánea:

a) Se realizan o repiten *n* ensayos Bernoulli.

b) El parámetro θ se mantiene constante entre ensayos.

c) Los ensayos son todos independientes entre sí.

Estas condiciones experimentales son muy frecuentes, y en general *el problema de interés* radica en el número de "éxitos" en *n* casos estudiados, o el número de respuestas "no" en *n* consultas, o el número de veces que ocurre un cierto fenómeno atmosférico en *n* observaciones realizadas.

Cuando se registra la ocurrencia de un fenómeno atmosférico en n observaciones suele utilizarse la distribución binomial para modelar el número total de ocurrencias. Sin embargo, en este caso es importante destacar que se debe verificar que las observaciones sean independientes y que la probabilidad de ocurrencia del fenómeno atmosférico (θ) se mantenga constante entre observaciones. En caso contrario el modelo binomial no será apropiado.

La falta de independencia entre observaciones en la agronomía es frecuente, y deberá tenerse en cuenta al momento de realizar un ensayo. La clave para modelar fenómenos en los que la independencia no puede asegurarse, está en reconocerla y luego incorporar esta información en la modelación. Si hay independencia entre las observaciones, entonces podemos seleccionar la distribución binomial. Más adelante se presentará una distribución que puede ser usada en algunos casos donde no hay independencia. Para ilustrar el concepto de independencia veamos por ejemplo en qué casos se puede presentar la falta de independencia en ensayos de germinación. Si se observa la germinación de semillas aisladas la respuesta de cada una de ellas no dependerá de lo que ocurrió en las otras. En este caso se registrarán *n* datos independientes. En cambio, si se realiza un ensayo de germinación en el que las semillas se encuentran en grupos (cajas de Petri) puede ocurrir que la no germinación de una semilla esté asociada a la presencia de hongos. Estos mismos pueden haber contaminado a las semillas vecinas y por lo tanto la respuesta de éstas no es independiente.

Se da a continuación una definición formal de distribución binomial.

Definición 3.6: Distribución Binomial.

Una variable aleatoria X tiene **distribución Binomial** si y sólo si su función de densidad es:

$$f(x;n,\theta)=\begin{cases}\binom{n}{x}\theta^{x}(1-\theta)^{n-x} & si\ x=0\ ,1,...,n\\ 0 & en\ caso\ contrario\end{cases}$$

donde $0\leq\theta\leq1$.

Las variables binomiales con parámetros n y θ se denotan como: **X**~ *Bin*(n,θ)

Nota: De forma análoga que en la distribución Bernoulli, f(x; n, θ) se caracteriza por dos parámetros: n y θ. Además, $\binom{n}{x}$ representa el número de combinaciones posibles de armar en base a n elementos en grupos de x, siendo $\binom{n}{x}=\dfrac{n!}{x!(n-x)!}$ y $n! = 1\times2\times...\times n$.

La E(**X)** cuando **X** tiene distribución Binomial se puede obtener a partir del siguiente desarrollo:

Como los posibles valores de x son 0, 1, 2,..., n, es posible escribir la esperanza como sigue:

$$\mu=E(X)=\Sigma_{x=0}^{n}xf(x)=\Sigma_{x=0}^{n}x\binom{n}{x}\theta^{x}(1-\theta)^{n-x}=\Sigma_{x=0}^{n}x\frac{n!}{x!(n-x)!}\theta^{x}(1-\theta)^{n-x}$$

Nótese que el primer valor de x es cero, y que si se reescribe $x!=x\ (x-1)!$ se puede simplificar la expresión anterior y quedar así:

$$=\Sigma_{x=1}^{n}x\frac{n!}{(x-1)!(n-x)!}\theta^{x}(1-\theta)^{n-x}$$

Luego, usando la misma técnica para $n!=n(n-1)!$ y observando que $\theta^{x}=\theta.\theta^{x-1}$, es posible sacar factor común $n\theta$, y reescribir la expresión anterior de la siguiente forma:

$$=n\theta\Sigma_{x=1}^{n}\binom{n-1}{x-1}\theta^{x-1}(1-\theta)^{n-x}$$

Si se hace el siguiente cambio de notación: $y = x - 1$ y $m = n - 1$, se tiene:

$$=n\theta\Sigma_{y=0}^{m}\binom{m}{y}\theta^{y}(1-\theta)^{m-y}=n\theta$$

ya que, $\Sigma_{y=0}^{m}\binom{m}{y}\theta^{y}(1-\theta)^{m-y}=1$ debido a que es la suma sobre todos los valores posibles de una función de probabilidad *Bin*(*m*,θ)= $\binom{m}{y}\theta^{y}(1-\theta)^{m-y}$

Si se calcula la varianza V(**X**) y siguiendo las ideas presentadas para el cálculo de la E(**X**), cuando **X** ~ *Bin*(*n*, θ) se verá que:

$$\sigma^2 = V(X) = n\theta(1-\theta)$$

Ejemplo 3.6

Supóngase que se toman 10 semillas de *Panicum maximum* Jacq. y se registra el evento “germinó” o “no germinó” después de 5 días desde su implantación. En este experimento las semillas están suficientemente aisladas como para asegurar respuestas independientes. Si la probabilidad de germinación es (para todas las semillas) igual a 0.25 calculemos:

a) Probabilidad que germinen 7 de las 10 semillas,

b) Probabilidad que germinen al menos 3 de las 10 semillas,

c) Probabilidad que germinen a lo sumo 5 semillas.

d) La esperanza de esta variable aleatoria.

e) La varianza.

Si X~ *Bin*(7; 10, 0.25), luego:

a) P(**X** = 7) = $\binom{10}{7}0.25^7\,(1\text{-}0.25)^{(10\,\text{-}7)} =$

$$\binom{10}{7}0.25^7(1-0.25)^{10-7} = \frac{10!}{7!(10-7)!}\;0.25^7\,0.75^3 = \frac{0.0185}{6} = 0.0031$$

b) P(**X** ≥ 3) = P(**X** = 3) + P(**X** = 4) + ... + P(**X** = 10) =
= 1 - (P(**X** = 0) + P(**X** = 1) + P(**X** = 2)) =
= 1 - (0.0563 + 0.1877 + 0.2816) = 0.4744

c) P(**X** ≤ 5) = P(**X** = 0) + P(**X** = 1) + ... + P(**X** = 5) =
= 0.0563 + 0.1877 + 0.2816 + 0.2503 + 0.1460 + 0.0584 = 0.9803

d) E(**X**) = 10 (0.25) = 2.5

e) V(**X**) = 10 (0.25) (1 - 0.25) = 1.875

Distribución Binomial Negativa

En conexión con la repetición de ensayos Bernoulli, ciertos *problemas de interés* centran su atención en “el número de ensayos necesarios hasta que ocurren *k* éxitos”.

Esta distribución también se la conoce como *distribución binomial para los tiempos de espera* o *distribución Pascal.*

Préstese atención a la siguiente secuencia:

a) Sea θ la probabilidad de éxito.

b) Se llama A al evento "el k-ésimo éxito ocurre en el ensayo número x".

c) Si el k-ésimo éxito ocurre en el ensayo x-ésimo, luego ya ocurrieron k-1 éxitos en los x-1 ensayos anteriores. Así, la probabilidad del evento B:"ocurren k -1 éxitos en x -1 ensayos" puede calcularse por la distribución binomial Bin(k-1; x-1, θ).

d) Si la probabilidad θ es constante entre los ensayos y C es el evento "éxito en el ensayo número x", luego P(C)=θ.

e) Entonces, P(A)=P(B$\cap$C). Como los eventos B y C son independientes, se tiene:

$P(A)= P(B\cap C)=P(B)*P(C)= \binom{x-1}{k-1}\ \theta^{k-1}\ (1 - \theta)^{x-k}\ )\ \theta = \binom{x-1}{k-1}\ \theta^{k}\ (1 - \theta)^{x-k}\ )$

Se da a continuación una definición formal de esta distribución.

Definición 3.7: Distribución Binomial Negativa (para k entero).

Una variable aleatoria **X** tiene **distribución Binomial Negativa** si y sólo si su función de densidad es:

$$f(x;k,\theta)=\begin{cases}\binom{x-1}{k-1}\theta^{k}\ (1-\theta)^{x-k} & \text{si } x = k, k+1, k+2...\\ 0 & \textit{en caso contrario}\end{cases}$$

donde $0\leq\theta\leq1$.

A las variables binomiales negativas con parámetros k y θ se las denotará como **X**~ *BinNeg*(k,θ).

<u>Nota</u>: De forma análoga a las otras distribuciones, f(x;k,θ) queda determinada por k y θ. Por otro lado, los valores de x son valores mayores o iguales que k. Obviamente no puede ocurrir k éxitos en *un número X de ensayos menor que k*.

Si se calcula la E(**X**), cuando **X**~*BinNeg*(k,θ) se verá que: $\mu = E(X) = \dfrac{k}{\theta}$

Si se calcula la varianza V(**X**) tendremos que: $\sigma^2 = V(X) = \dfrac{k}{\theta}\left(\dfrac{1}{\theta}-1\right)$

Desde el punto de vista del cálculo de probabilidades, en problemas específicos, no es necesario acudir a una tabla de probabilidades de binomial negativa, si se dispone de una tabla binomial. La relación entre una y otra distribución es la siguiente:

$$BinNeg(x;\, k,\, \theta) = \frac{k}{x} \,.\, Bin(k;\, x,\, \theta)$$

Ejemplo 3.7

Un acopiador de granos recibe camiones cargados con maíz. La carga puede venir con o sin semillas de chamico. La probabilidad de que el camión venga "limpio" es 0.90. Si el silo se llena con 20 camiones y se desea que sean solo con cargas "limpias".

a) ¿Cuál es la probabilidad de que los llene con los primeros 20 camiones?

b) ¿Cuál es la probabilidad de que los llene con los primeros 30 camiones?

c) ¿Cuál es el número de camiones necesarios para llenar el silo con carga limpia con probabilidad 0.95.

a) Tomando $x = 20$, $k = 20$ y $\theta = 0.90$, se tiene:

$$BinNeg(20;20;0.9) = \frac{k}{x} \; Bin(k;\, x,\theta) = \frac{20}{20} \,.\, Bin(20;\, 20,\, 0.9) = 0.1216$$

b) Tomando $x = 30$, $k = 20$ y $\theta = 0.90$, se tiene:

$$BinNeg(30;20;0.9) = \frac{k}{x} \,.\, Bin(k;\, x,\, \theta) = \frac{20}{30} \times Bin(30;\, 20,\, 0.9) = 0.00024$$

c) Para contestar esta pregunta, se debe calcular la probabilidad de que los 20 camiones "limpios" se hayan conseguido con X camiones o menos. Obviamente X no puede ser menor que 20. Luego, cuando la probabilidad de que se hayan conseguido los 20 camiones limpios con X camiones o menos, alcance (o esté cerca de) 0.95, se tendrá la solución. La siguiente tabla muestra para distintos números totales de camiones (X), la probabilidad de que se alcancen 20 "éxitos" en el X-ésimo (llamando "A" a este evento y P(A) a su probabilidad) y también la probabilidad de que estos éxitos se alcancen en el X-ésimo camión o en alguno anterior (evento B).

Camiones	P(A)	P(B)
20	0.1216	0.1216
21	0.2432	0.3647
22	0.2553	0.6200
23	0.1872	0.8073
24	0.1077	0.9149
25	**0.0517**	**0.9666**
26	0.0215	0.9881
27	0.0080	0.9961
28	0.0027	0.9988
29	0.0008	0.9997
30	0.0002	0.9999

Teniendo en cuenta la información anterior, se puede concluir que con 25 camiones, es altamente probable (P=0.9666) que se complete el silo con carga "limpia".

Existen muchas aplicaciones de la distribución binomial negativa cuando k=1. Por esto recibe el nombre especial de **distribución geométrica**, cuya definición se da a continuación.

Distribución Geométrica

Definición 3.8: Distribución Geométrica.

Una variable aleatoria **X** tiene **distribución Geométrica** si y sólo si su función de densidad es:

$$f(x,\theta) = \begin{cases} \theta \quad (1\text{-}\theta)^{\,x-1} & \text{si } x = 1,2,3... \\ 0 & \textit{en caso contrario} \end{cases}$$

donde $0 \le \theta \le 1$.

Esta distribución tiene especial aplicación cuando el problema de interés es modelar la distribución del número de ensayos necesarios para encontrar el primer éxito.

Ejemplo 3.8

El mismo acopiador del ejemplo anterior tiene problemas financieros, por eso decide vender a razón de un camión de grano por día.

Sabiendo que la probabilidad diaria de que un productor reclame la venta de un camión es de 0.2 y esa probabilidad se mantiene constante durante el próximo mes, ¿cuál es la probabilidad de vender un silo sin que le reclamen ninguna venta?

Esta situación implica que lo que pretende el acopiador es vender todo un silo (20 camiones) sin que durante ese período tenga que desembolsar un pago.

Lo que este acopiador desea calcular es la probabilidad de que el primer reclamo ocurra en el día 21 o posteriormente.

Esta probabilidad debería calcularse como

$P(X=21) + P(X=22)+....=1- (P(X=20) + P(X=19) +...+ P(X=1))$

donde X~Geom(x,0.2). Luego, utilizando la definición y las propiedades de las funciones de distribución tenemos:

$1-(P(X=20) + P(X=19) +...+ P(X=1)) =$

$=1-[0.2(1-0.2)^{(20-1)} + 0.2(1-0.2)^{(19-1)} +...+ 0.2(1-0.2)^{(1-1)}] = 0.0115$

En consecuencia, es muy poco probable que el acopiador pueda vender un silo a razón de un camión diario sin que ningún productor le pida una venta durante ese período.

Distribución Hipergeométrica

Esta distribución está ligada a situaciones de **muestreo sin reposición**, es decir situaciones en que al azar se elige un elemento de una población y así sucesivamente hasta completar la muestra, sin restituir los elementos extraídos.

Para inducir la fórmula de esta distribución, análoga a la binomial, considérese como población a un conjunto de *N* elementos de los cuales *k* poseen uno de dos estados posibles (éxito) y *N-k* que presentan el otro (fracaso).

Al igual que en la binomial, el *problema de interés* es "hallar la probabilidad de obtener X éxitos, pero en este caso, cuando se seleccionan sin reposición *n* elementos de un conjunto de *N* ".

Como se recordará, el concepto frecuencial de probabilidad está asociado al cociente:

$$\frac{\text{número de casos favorables}}{\text{número de casos posibles}}$$

En este problema el *número de casos totales* viene dado por el número de combinaciones posibles que se puede obtener a partir de *N* elementos tomados de a grupos de *n*. Esto es:

$$\text{Número de casos totales} = \binom{N}{n}$$

El *número de casos favorables* vendrá dado por el número de formas posibles de elegir *x* éxitos y *n-x* fracasos del conjunto de *N* elementos en los que hay *k* éxitos y *N-k* fracasos, por lo que este número será el siguiente producto:

$$\text{Número de casos favorables} = \binom{k}{x}\binom{N-k}{n-x}$$

donde $\binom{k}{x}$ indica el numero de formas posibles en las que se pueden escoger "x" éxitos de un conjunto de "k" éxitos y análogamente $\binom{N-k}{n-x}$ indica el número de formas posibles en las que se pueden escoger "n-x" fracasos de un total de "N-k" fracasos.

Luego para cada forma de elegir un conjunto de "x" éxitos existen $\binom{N-k}{n-x}$ formas de obtener "n-x" fracasos y de allí el producto.

Se da a continuación una definición formal de esta distribución.

Definición 3.9: Distribución Hipergeométrica.

Una variable aleatoria **X** tiene **distribución Hipergeométrica** si y sólo si su función de densidad es:

$$f(x;n,N,k) = \begin{cases} \dfrac{\binom{k}{x}\binom{N-k}{n-x}}{\binom{N}{n}} & si\ x = 0\ ,1,...,n;\ x \le k;\ n-x \le N-k \\ 0\ en\ caso\ contrario. \end{cases}$$

Nota: De forma análoga que en las otras distribuciones, esta función de densidad posee tres parámetros: *n, N, k*.

Se denotará a las variables hipergeométricas con parámetros n, N, k con la siguiente expresión: **X**~*Hiper*(*n,N,k*)

Si se calcula la E(**X**) cuando **X** tiene distribución hipergeométrica, se verá que:

$$\mathrm{E(X)} = \frac{\mathrm{n\ k}}{\mathrm{N}}$$

Si se calcula la varianza V(**X**) se verá que:

$$\sigma^2 = V(X) = \frac{n\ k\ (N-k)\ (N-n)}{N^2\ (N-1)}$$

Ejemplo 3.9

Cuando la semilla de maíz viene contaminada con chamico, el precio de esta semilla es inferior. Para determinar el precio que debe pagar por un determinado lote, un Ingeniero Agrónomo decide examinar 20 de 500 bolsas de semillas de maíz. Si el 10% de las bolsas (50) contienen semillas de chamico, ¿Cuál es la probabilidad de que ninguna de ellas esté incluida en la muestra?

Tomando x=0, n=20, N=500 y k=50 en la función hipergeométrica, se tiene:

$$Hiper(0;20,500,50)=\frac{\binom{50}{0}\binom{450}{20}}{\binom{500}{20}}=0.1164$$

Vale decir que de cada 100 veces que el ingeniero realiza esta prueba, que consiste en tomar una muestra de 20 bolsas de un total de 500 donde al menos 50 están contaminadas, en el 88% de las veces (al menos) encontrará bolsas con chamico.

En oportunidad de presentar la distribución binomial había quedado planteado el hecho de que cuando *N* es grande, el considerar si el muestreo es con o sin reposición puede ser insignificante. Considérese el siguiente ejemplo:

Ejemplo 3.10

Entre 120 cámaras de germinación, 80 están bien calibradas. Si se toma una muestra aleatoria de 5 cámaras, hallar la probabilidad de que solamente 2 de las 5 estén bien calibradas en base a:

a) La distribución hipergeométrica

Tomando x=2, n=5, N=120 y k=80, se tiene:

$$Hiper(2;\,5,\,120,\,80)=\frac{\binom{80}{2}\binom{40}{3}}{\binom{120}{5}}\approx 0.164$$

b) La distribución binomial.

Tomando x=2, n=5, $\theta=\frac{80}{120}=\frac{2}{3}$ en la fórmula de la binomial:

$$Bin(2;5,2/3) = \binom{5}{2}(2/3)^2\ (1\text{-}2/3)^3 \approx 0.165$$

Distribución Poisson

La distribución de Poisson da un modelo para variables de tipo conteo, donde los conteos se refieren al registro del número de un evento de interés en una unidad de tiempo o espacio dados (horas, minutos, m^2, m^3, etc.).

Ejemplos de variables que se pueden modelar como Poisson son:

a) Número de huevos de un insecto en una oviposición.

b) Número de bacterias en una muestra de agua.

c) Número de semillas defectuosas observadas en una cinta transportadora por minuto.

d) Número de nemátodos por unidad de volumen del suelo.

e) Número de pulgones por planta.

f) Número de pulgones por m^2.

Se da ahora la siguiente definición formal para esta distribución.

Definición 3.10: Distribución Poisson.

Una variable aleatoria **X** tiene **distribución Poisson** si y sólo si su función de densidad es:

$$f(x,\lambda) = \begin{cases} \dfrac{\lambda^x e^{-\lambda}}{x!} & si\ x = 0,1,2,... \\ 0\ en\ caso\ contrario \end{cases}$$

Nota: Se indica que X tiene distribución de Poisson con parámetro λ, con la siguiente notación: **X**~*Poisson*(λ).

Si se calcula la E(**X**) y la V(**X**), cuando **X**~ *Poisson*(λ), se obtiene:

$$\mu = E(\mathbf{X}) = \lambda$$

$$\sigma^2 = V(X) = \lambda$$

En esta distribución la varianza es igual a la esperanza y por lo tanto la variabilidad de los conteos aumenta con el nivel medio de los mismos. Este es un caso típico de asociación entre esperanza y varianza.

Ejemplo 3.11

Si el número promedio de picaduras de gorgojo por semilla es 0.2 (es decir, por ejemplo que, en promedio, cada 100 semillas se cuentan 20 picaduras), ¿cuántas de 100 semillas no tendrán picaduras?, ¿cuántas 1 picadura? y ¿cuántas 2 o más?

Para responder a este problema se calcula la probabilidad de que una semilla tomada al azar tenga una picadura o ninguna picadura, suponiendo distribución Poisson para esta variable. Luego:

$$P(X=0)= \frac{0.2^0 e^{-0.2}}{0!} = 0.819$$

$$P(X=1)= \frac{0.2^1 e^{-0.2}}{1!} = 0.164$$

y $P(X>1)= 1-[P(X=0)+P(X=1)]=1-0.982= 0.018$

En consecuencia, si la probabilidad de que una semilla tomada al azar no tenga picaduras es 0.819, deberíamos esperar que, en un grupo de 100, aproximadamente 82 no estén picadas, y si la probabilidad de que tengan solo una picadura es de 0.164, entonces solo 16 semillas cumplirán esta condición y finalmente, aproximadamente 2 de cada 100 semillas tendrán 2 o más picaduras.

Nota: Existe una relación entre la distribución de Poisson y la Binomial que permite aproximar las probabilidades de variables binomiales cuando n es grande y θ pequeño. En estos casos se puede tomar $\lambda= n\theta$ y calcular las probabilidades de éxito bajo esta distribución.

Distribución Multinomial

Esta distribución puede ser vista como una generalización de la distribución binomial, donde el interés es calcular la probabilidad de obtener $n_1, n_2, ..., n_k$ en k categorías en una muestra de tamaño $N=n_1+n_2+...+n_k$ conociendo que la probabilidad de ocurrencia de cada categoría en la población $\theta_1, \theta_2,...., \theta_k$. Se dice entonces que una variable tiene distribución multinomial y se denota como *Multi*$(N,\theta_1, \theta_2,...., \theta_{k-1})$, cuando su función de densidad está dada por:

Definición 3.11: Distribución Multinomial.

Si X_1, X_2,..., X_k representan las ocurrencias de las K categorías en la población entonces decimos que (X_1, X_2,..., X_k) tiene **distribución Multinomial** si su densidad es la siguiente:

$$f(x_1, x_2, ..., x_k; \theta_1, \theta_2, ..., \theta_k) = \begin{cases} \dfrac{N!}{x_1!.x_2!....x_k!}\theta_1^{x_1}\theta_2^{x_2}...\theta_k^{x_k} \\ 0 \quad en \quad caso \quad contrario \end{cases}$$

donde $N=x_1+x_2+...+x_k$; $x_i \in [0,1,..N]$; $0 \le \theta_i \le 1$, $i=1,...,k$; $\theta_1+\theta_2+...+\theta_k=1$.

Ejemplo 3.12

En un cultivo el ataque de una enfermedad puede ser calificado como severo, moderado o sin ataque. Supóngase que la probabilidad de ataque severo es de 0.05 y de moderado de 0.20. ¿Cuál es la probabilidad de encontrar sobre un total de 10 plantas observadas, 2 con ataque severo, 2 con ataque moderado y 6 sanas?

Solución: $P(X_1=2, X_2=2, X_3=6) = \dfrac{10!}{2!2!6!} 0.05^2 \ 0.20^2 \ 0.75^6 = 0.0224$

Funciones de densidad de variables aleatorias continuas

A continuación se presentan algunas funciones de distribución continuas que aparecen frecuentemente en las aplicaciones prácticas.

Distribución Uniforme

Así como en la sección anterior se definió la distribución uniforme para variables aleatorias discretas, ahora se presenta una distribución análoga para el caso continuo. Su definición es la siguiente:

Definición 3.12: Distribución Uniforme

Una variable aleatoria **X** tiene **distribución Uniforme** si y sólo si su función de densidad es:

$$f(x) = \begin{cases} \dfrac{1}{\beta - \alpha} si \ \alpha < x < \beta \\ 0 \ en \ caso \ contrario \end{cases}$$

Nota: Cuando una variable aleatoria tiene distribución uniforme con parámetros α y β, se indica como **X~**U(α,β).

La E(**X**) y la V(**X**) cuando **X** tiene densidad uniforme son:

E(**X**) = $(\alpha + \beta) / 2$ V(**X**) = $(\beta - \alpha)^2 / 12$

Una de las principales aplicaciones de esta distribución es en estudios de simulación Montecarlo, ya que a partir de esta función es posible generar números pseudoaleatorios de otras distribuciones. En todos los lenguajes de programación o incluso en las planillas de cálculo existen declaraciones (a modo de funciones o procedimientos) para generar números con distribución uniforme.

Los números generados por computadoras se dicen que son pseudoaleatorios, y no aleatorios, ya que el mecanismo que los genera es *determinístico*. Ocurre que el algoritmo que se elige para generar un número uniforme simula o aparenta ofrecer números tomados al azar. Mientras mejor simule la producción de números aleatorios tanto mejor el algoritmo. En la mayoría de las nuevas revisiones de los lenguajes de programación se ofrecen buenos generadores de números uniformes pseudoaleatorios.

Distribución Gamma

La función de distribución Gamma es importante en estadística ya que hay otras distribuciones de uso frecuente (exponencial y chi-cuadrado) que son casos particulares de ella, y que juegan un rol fundamental en variados campos de las aplicaciones y la teoría estadística. Asimismo, en el campo de las Ciencias Agropecuarias, esta distribución cobra importancia en la modelación de algunos fenómenos meteorológicos como las precipitaciones.

A continuación se da una definición de esta distribución.

Definición 3.13: Distribución Gamma.

Una variable aleatoria **X** tiene **distribución Gamma** si y solo si su función de densidad es:

$$\text{f}(x) = f(x) = \begin{cases} \dfrac{x^{a\text{-}1}\ e^{\text{-}x/b}}{\beta^{\alpha}\Gamma(\alpha)} & si\ 0 \le x \le \infty \\ 0\ en\ caso\ contrario \end{cases}$$

con α>0, β>0 y

donde: $\Gamma(\alpha) = \int_0^{\infty} y^{\alpha-1} e^{-y}\, dy$ es conocida en matemáticas como función gamma.

Nota: Se indica que una variable aleatoria X tiene una distribución Gamma con: **X**~*G*(α,β).

La función de densidad de una distribución *gamma* es una función asimétrica, que tiende a la simetría para ciertos valores de sus dos parámetros. A modo de ejemplo se presenta en la siguiente gráfica formas diferentes de la densidad Gamma.

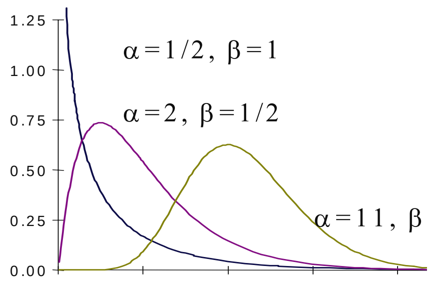


Figura 3.7: Forma de la densidad Gamma, para distintos valores de sus parámetros

El máximo (si existe) en esta función de densidad viene dado por $x = \beta(\alpha - 1)$

Por último, es posible mostrar que: $E(\mathbf{X}) = \alpha\beta$; $V(\mathbf{X}) = \alpha\beta^2$

Distribución Exponencial

Esta densidad es un caso especial de la función de densidad $G(\alpha,\beta)$, tomando α=1 y $\beta=\theta$, quedando así definida:

Definición 3.14: Distribución Exponencial.

Una variable aleatoria **X** tiene **distribución Exponencial** si y sólo si su función de densidad es:

$$f(x) = \begin{cases} \dfrac{e^{-x/\theta}}{\theta} & si\ 0 \le x \\ 0 & en\ caso\ contrario \end{cases}$$

donde $\theta > 0$.

Nota: En la práctica denotaremos que una variable aleatoria tiene distribución exponencial con parámetro θ con la siguiente notación: **X** ~ *Exp*(θ).

Es posible mostrar que: $E(\mathbf{X}) = \theta$; $V(\mathbf{X}) = \theta^2$

La distribución exponencial también es conocida como la distribución de los tiempos de espera y es utilizada para calcular la probabilidad de que un instrumento electrónico falle, pasado un cierto tiempo, o el tiempo necesario para que ocurra un accidente de transito en una ruta con probabilidad 0.90, etc.

Distribución Chi-Cuadrado

La distribución chi-cuadrado aparece con mucha frecuencia en la estadística aplicada ya que los llamados “test” del cociente de máxima verosimilitud que se utilizan en la prueba de hipótesis estadísticas tienen todos distribución asintótica chi-cuadrado. Cuando se construye un “test”con las técnicas de la máxima verosimilitud, se usa esta distribución, al menos para muestras grandes.

A modo de ejemplo, las técnicas de bondad de ajuste que se usan en genética para establecer si una frecuencia fenotípica se ajusta a un modelo de herencia mendeliana, utilizan un estadístico cuya distribución (asintótica) es chi-cuadrado y de allí su nombre (“test” de chi-cuadrado).

Esta distribución también aparece relacionada a la distribución de la varianza muestral, que estudiaremos más adelante. Por ahora sólo daremos su definición formal.

Definición 3.15: Distribución Chi-Cuadrado

Una variable aleatoria **X** tiene **distribución Chi-Cuadrado** si y sólo si su función de densidad es:

$$f(x) = \begin{cases} \dfrac{x^{(\nu-2)/2} e^{-x/2}}{2^{\nu/2}\Gamma(\nu/2)} \, si \ 0 \le x \\ 0 \ en \ caso \ contrario \end{cases}$$

donde ν es un entero positivo conocido como grados de libertad de la distribución.

Nota: Se denotará a las variables chi-cuadrado con ν grados de libertad como: X~χ^2_ν , siendo ν el único parámetro de esta distribución.

Es posible mostrar que: E(**X**) = ν ; V(**X**) = 2ν

Obsérvese que la esperanza es igual a los grados de libertad de la distribución y que la varianza es también una función lineal de este parámetro.

Una forma alternativa de definir variables aleatorias chi-cuadrado es a partir de variables aleatorias normales estándar, como de muestra a continuación:

Sean X_1, X_2,....,X_ν variables aleatorias normales independientes con esperanza $\mu = 0$ y varianza $\sigma^2 = 1$. Luego si $Y = \sum_{i=1}^{\upsilon} X_i^2$,

Entonces Y tiene **distribución** χ^2 con ν grados de libertad.

Esta distribución esta involucrada en la definición de las distribuciones T de Student y F que se presentarán en el Capítulo 4.

Ejercicios

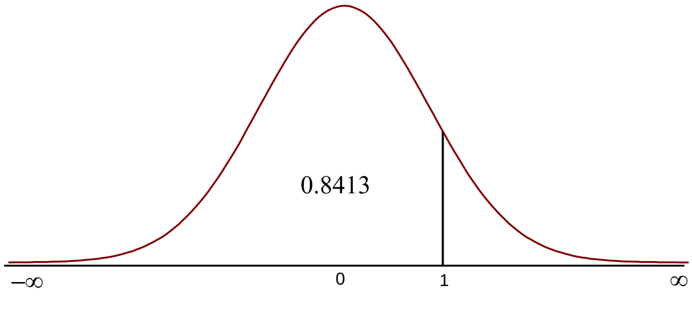
Ejercicio 3.1: Uso de la tabla de cuantiles de la Distribución Normal Estándar

Esta tabla presenta 2 columnas:

La primera columna se refiere a la distancia desde un valor a la media medida en número de desviaciones típicas (valores de la variable Z). Por ejemplo el valor 1 en esta columna indica una desviación estándar por encima de la media y -1.7 corresponde a 1.7 desviaciones estándar por debajo de la media.

La segunda columna contiene el área bajo la curva normal entre $-\infty$ y el valor correspondiente a la primer columna, es decir el valor de la función de distribución normal estándar acumulada. Por ejemplo para el valor 1 de z, el área asociada es 0.8413.

Así se puede concluir que la probabilidad de que una variable distribuida normalmente con $\mu = 0$ y $\sigma^2 = 1$ tome valores iguales o menores que 1, es igual a 0.8413 (1 es por lo tanto el cuantil 0.8413 de la distribución normal estándar), lo que se ilustra en la siguiente figura:



Usando la tabla de cuantiles de la Distribución Normal Estándar obtener las siguientes probabilidades:

a) $P(Z \leq 1.3)$ b) $P(Z \leq 4)$ c) $P(Z \geq 1.3)$
d) $P(-1 \leq Z \leq 1)$ e) $P(0.5 \leq Z \leq 1)$ f) $P(Z = 1)$

Ejercicio 3.2

Por medio de un tamiz de malla de 8 mm de diámetro se zarandean 8000 granos de maíz. El *diámetro del grano* de maíz sigue una distribución normal con esperanza igual a 9 mm y una desviación estándar de 1.2 mm.

a) ¿Qué proporción de granos serán retenidos por el tamiz?

b) ¿ Qué proporción de granos no retenidos, serán retenidos por un tamiz de diámetro de malla igual a 7.5 mm?

c) ¿Qué proporción de granos pasará a través de los dos tamices?

Ejercicio 3.3

Si X es una variable aleatoria distribuida normalmente con $\mu = 10$ y $\sigma^2 = 4$.

a) ¿Cuál es la probabilidad de que X tome valores menores que 9?

b) ¿Cuál es la probabilidad de que X tome valores entre 9 y 11?

Ejercicio 3.4

La variable *altura de plántulas* para una población dada se distribuye normalmente con media $\mu = 170$ mm y $\sigma = 5$ mm. Encontrar la probabilidad de los siguientes eventos:

a) Plantas con alturas de al menos 160 mm.

b) Plantas con alturas entre 165 y 175 mm.

Ejercicio 3.5

Si la variable *espesor de un sedimento* en un sustrato de suelo, se distribuye normalmente con media $\mu = 15$ micrones y desviación estándar $\sigma = 3$ micrones.

a) ¿Cuál es el cuantil 0.75 de la distribución de la variable?

b) ¿Cómo se interpreta este valor?

Ejercicio 3.6

La *altura de plantas de soja* de la variedad Hood se distribuye aproximadamente normal con media 55 cm y desviación estándar de 5.8 cm. Por otro lado, la *altura de plantas de yuyo colorado* (Amaranthus sp.) invasora de este cultivo, también se distribuye en forma normal con media 62 cm y desviación estándar de 3 cm. Si se decide aplicar un herbicida usando un equipo a sogas:

a) ¿A qué altura debe disponerse la soga para eliminar el 90% de la maleza en este cultivo?

b) ¿Suponiendo que el herbicida no es selectivo, es decir mata por igual a toda planta que toma contacto con la soga, ¿qué porcentaje de plantas de soja se perderá a la altura de soga encontrada en el punto anterior?

Ejercicio 3.7

El caudal de un *canal* de riego medido en m^3/seg es una variable aleatoria con distribución aproximadamente normal con media 3 m^3/seg. y desviación estándar 0.8 m^3/seg. A partir de estas referencias calcular la probabilidad de los siguientes eventos:

a) Evento A: que el caudal en un instante dado sea a lo sumo de 2.4 m^3/seg.

b) Evento B: que el caudal en un instante dado esté entre 2.8 y 3.4 m^3/seg.

Ejercicio 3.8

Una empresa exportadora de manzanas necesita encargar 10000 cajones para el embalaje de la fruta. Sin embargo, no todos los cajones son iguales ya que sus especificaciones dependen de la calidad del producto envasado. Así, de acuerdo al diámetro de la manzana se identifican 3 categorías de calidad.

Categoría I: manzanas cuyo diámetro es menor de 5 cm

Categoría II: manzanas cuyo diámetro está comprendido entre 5 y 7 cm

Categoría III: manzanas cuyo diámetro es mayor que 7 cm

Las frutas de mayor calidad son las correspondientes a la categoría II por su tamaño y homogeneidad. Si la distribución del diámetro de las manzanas puede modelarse bien mediante una distribución normal con media $\mu = 6.3$ y varianza $\sigma^2 = 2$, responder:

¿Cuántos cajones se necesitarán para cada categoría de manzanas?

Ejercicio 3.9

Siguiendo con el ejercicio anterior y conociendo el comportamiento cíclico de la demanda de cada categoría de manzanas, se sabe que en la presente campaña va a tener más demanda la manzana de la categoría II (manzanas con diámetro entre 5 y 7 cm), con lo cual las ganancias para el exportador se maximizarían en caso de aumentar el volumen de la cosecha para esta categoría. Una forma de regular el tamaño final de esta fruta es mediante la eliminación temprana de los frutos en formación (raleo). Si se eliminan muchos frutos el tamaño final de las manzanas será mayor que si se eliminan pocos o ninguno.

La experiencia ha permitido establecer las características distribucionales del *diámetro final de las manzanas* bajo dos estrategias de manejo:

A: no eliminar ningún fruto

B: eliminar 1 de cada 3 manzanas

La estrategia A produce frutos con diámetros distribuidos N (6.3, 2.0) y la estrategia B produce frutos con diámetros distribuidos N (6.8, 0.9).

¿Cuál de las dos estrategias produce mayor proporción de frutos de Categoría II?

Ejercicio 3.10

El *espesor de la cáscara del huevo* determina la probabilidad de ruptura desde que la gallina lo pone hasta que llega al consumidor. El espesor, medido en centésimas de milímetro, se distribuye normal y se sabe que:

a) se rompen el 50 % de los huevos con espesor de cáscara menor a 10 centésimas de mm (cmm).
b) se rompen el 10 % de los huevos cuyo espesor de cáscara está comprendido entre 10 y 30 cmm.
c) no se rompen los huevos con espesor de cáscara mayor de 30 cmm.

Si en un establecimiento avícola la media del espesor de cáscara es de 20 cmm y la desviación estándar de 4 cmm:

¿Cuántos, de los 5000 huevos que se producen diariamente, llegan sanos al consumidor?

Ejercicio 3.11

El día de floración de una hortaliza (en escala juliana:1-365 días) se puede modelar con una distribución normal centrada en el 18 de agosto (día 230) y con desviación estándar de 10 días. Si desde la fecha de la floración hasta la cosecha hay un lapso de 25 días:

a) ¿Qué proporción de la cosecha se habrá realizado para el 16 de septiembre (día 259)?

b) Si se considera primicia a los frutos obtenidos antes del 1 de septiembre (día 244): ¿qué proporción de la cosecha se espera que sea primicia?

c) Si la ganancia es de 2 pesos por cajón y se espera una producción total de 1500 cajones, ¿cuál es la ganancia esperada con los cajones primicia, son un 30% más caros?

d) La aplicación de un regulador del crecimiento permite adelantar 3 días la fecha de floración y reduce la desviación estándar de 10 a 6 días. Si la ganancia por cajón se reduce en 5 centavos debido al costo del regulador: ¿produce su aplicación un aumento del porcentaje de frutos primicia?

Ejercicio 3.12

Un fitomejorador desea controlar la variabilidad de los brotes comerciales de espárrago, ya que las normas de embalaje establecen una longitud máxima de cajas de 23.5 cm. Suponiendo que la longitud de los brotes de este cultivo se distribuye normalmente, con una esperanza igual a 21 cm. ¿Cuál debería ser el valor de la desviación estándar del carácter longitud del brote, para que la probabilidad de que existan espárragos que no puedan ser embalados, no sea mayor a 0.05?

Ejercicio 3.13

Un Ingeniero Agrónomo del Servicio de Alerta contra Fitóftora de una región viñatera afirma que 2 de cada 10 lotes afectados por la enfermedad se deben al mal manejo de los mismos. ¿Cuál es la probabilidad que:

a) de 100 lotes, a lo sumo 10, sean afectados por la enfermedad, por problemas de mal manejo?

b) de 100 lotes, ninguno presente la enfermedad por problemas de mal manejo?

Ejercicio 3.14

Un Ingeniero especialista en control de calidad de semillas de trigo, afirma que la empresa para la cual trabaja, produce un 95% de las bolsas de semilla de trigo con una pureza del 99%. Si fuera cierta su afirmación, ¿cuál sería la probabilidad que:

a) de 20 bolsas tomadas al azar, todas satisfagan que no poseen más del 1% de cuerpos extraños?

b) de 20 bolsas tomadas al azar, a menos 2 posean más del 1% de cuerpos extraños.

Ejercicio 3.15

Si la probabilidad de que un productor adopte una técnica, divulgada por un Instituto de Investigación Agropecuaria, es de 0.75, hallar la probabilidad que:

a) el décimo productor en tener acceso a la documentación de divulgación sea el primero en adoptarla.

b) el décimo productor en tener acceso a la documentación de divulgación sea el quinto en adoptarla.

Ejercicio 3.16

Se quiere encontrar plantas de trigo con propiedades resistentes a los pulgones. Un síntoma de resistencia es la ausencia de pulgones en la planta. Se calcula que la frecuencia de plantas sin pulgones en un cultivo es de alrededor de 1/200 pero solo 1 de cada 10 de estas plantas presentan genes de resistencia. ¿cuántas plantas de trigo deberán revisarse para tener una probabilidad de al menos 0.95 de encontrar una con los genes de resistencia?

Ejercicio 3.17

Un técnico en semillas desea inspeccionar el funcionamiento de 20 cámaras de cría. Para esto toma dos cámaras al azar y registra la temperatura de las mismas. Si estas dos cámaras funcionan correctamente, el grupo de 20 será aceptado. Cuáles son las probabilidades que tal grupo de 20 cámaras sea aceptado si contiene:

a) 4 cámaras con registros de temperaturas no adecuadas;

b) 8 cámaras con registros de temperatura no adecuadas;

c) 12 cámaras con registros de temperaturas no adecuadas.

Ejercicio 3.18

En una red de computadores asociados a estaciones agroclimatológicas y dedicadas a transmitir la información registrada a un computador central (servidor) vía telefónica, el 1.4% de los llamados desde los computadores al servidor dan ocupado. Determinar las probabilidades de que de 150 intentos de comunicaciones (llamados) sólo en 2 casos de ocupado el servidor.

Ejercicio 3.19

En un experimento, el error cometido en determinar la densidad de una substancia es una variable aleatoria con distribución uniforme, con α= -0.015 y β= 0.015. Hallar las probabilidades que:

a) El error esté entre 0.01 y 0.02;

b) El error exceda 0.005.

Ejercicio 3.20

Un Investigador ha establecido como hipótesis de trabajo, en base a experiencias previas bajo condiciones controladas, que la producción de oxígeno durante la fotosíntesis de la alfalfa sigue una distribución Gamma(3,2). ¿Cuál será la producción promedio y la varianza con estos parámetros?

4

Distribución de Estadísticos Muestrales

Introducción

El objetivo del muestreo es inferir propiedades de una población a partir de una fracción de ella, conocida como muestra. Desde el punto de vista estadístico, lo que se pretende conocer son los parámetros de la distribución de la variable de interés. El muestreo tiene por objeto proveer información esa distribución. Luego, los estadísticos muestrales sirven como aproximación (estimación) de los parámetros que caracterizan a la distribución. Por otra parte, los estadísticos son variables aleatorias y como tales, tienen una distribución asociada. Los objetivos de este capítulo son: comprender la naturaleza aleatoria de los estadísticos muestrales, estudiar las propiedades estadísticas de la media y varianza muestrales y adquirir destrezas en el cálculo de probabilidades asociadas a estos estadísticos.

Distribución del estadístico media muestral

Dado que la media muestral es una variable aleatoria (note que su valor varía de muestra a nuestra), nos interesa conocer su distribución. Cuando se estudian las distribuciones de los estadísticos muestrales se hace desde un punto de vista teórico, suponiendo poblaciones de tamaño infinito. Si se quieren observar estas propiedades partiendo de poblaciones finitas, a través de simulación, se recurre a la técnica de muestreo con reposición[3] porque de esa forma se emula una población de tamaño infinito.

Ejemplo 4.1

Considérese por ejemplo la distribución de una variable aleatoria definida sobre un espacio muestral y la distribución de la media muestral obtenida por muestreo aleatorio simple con reposición para muestras de tamaño 2. Para ello suponga una

---

3 Se entiende por muestreo aleatorio con reposición a aquel donde las unidades seleccionadas pueden repetirse dentro de la muestra y entre muestras.

población (finita) de cuatro plantas de zapallos (N = 4) donde la característica de interés es el número de zapallos por planta. Luego si se toma una planta al azar y se observa el número de frutos, se puede homologar el resultado de este experimento a una variable aleatoria discreta (X). Los valores de la variable X en la población y su función de densidad se presentan en la Tabla 4.1; mientras que la Figura 4.1 representa gráficamente la función de densidad.

Tabla 4.1: Función de densidad del número de frutos en una población de 4 plantas de zapallo

Planta	X = Nº de Frutos	$f(x_i)$
P_1	3	1/4
P_2	2	1/4
P_3	1	1/4
P_4	4	1/4

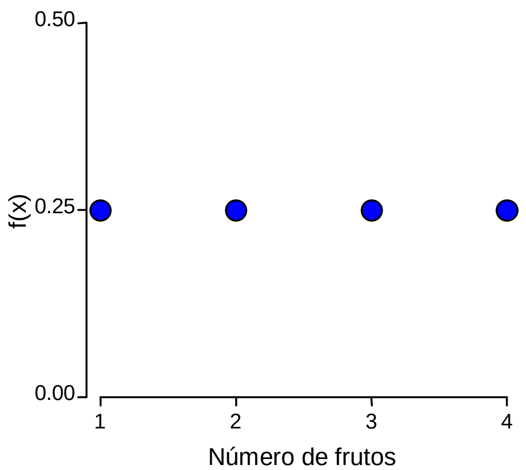


Figura 4.1: Función de densidad de X = número de frutos

Nota: Este tipo de función con idéntica densidad para todos los valores de x, se conoce como densidad uniforme.

De acuerdo a las definiciones de esperanza y varianza para variables discretas, dadas en el Capítulo 2, se tendrá:

$$\mu = \sum_i x_i f(x_i)$$

$$\mu = 1\,\frac{1}{4} + 2\,\frac{1}{4} + 3\,\frac{1}{4} + 4\,\frac{1}{4} = \frac{1+2+3+4}{4} = 2.5$$

$$\sigma^2 = \sum_i \left(x_i - \mu\right)^2 f(x_i)$$

$$\sigma^2 = (1\text{-}2.5)^2\,\frac{1}{4} + (2\text{-}2.5)^2\,\frac{1}{4} + (3\text{-}2.5)^2\,\frac{1}{4} + (4\text{-}2.5)^2\,\frac{1}{4} = 1.25$$

Tomando muestras de dos plantas con reposición, hay N^2 muestras posibles para extraer, esto es 4^2=16 muestras. Este es un espacio muestral finito que tiene 16 resultados posibles, todos con igual probabilidad. Si a cada resultado posible del muestreo se le asocia un valor correspondiente al promedio del número de frutos de las plantas obtenidas en la muestra, se obtiene una variable aleatoria llamada **media muestral basada en muestras de tamaño *n* = 2.**

La tabla 4.2 presenta todos los posibles resultados del proceso y el valor de la variable aleatoria media muestral, basada en muestras de tamaño n = 2.

Si bien todos los resultados posibles tienen igual probabilidad, en términos de la variable aleatoria "media muestral", varios de estos producen el mismo resultado. Por lo tanto un valor de media muestral "reúne" varios resultados elementales en un único evento.

Por ejemplo $\overline{X}$ = 3, corresponde al evento A={P_1P_1, P_2P_4 ,P_4P_2}, luego aplicando los axiomas de probabilidad: **P($\overline{X}$ = 3)** = P(A) = P(P_1P_1) + P(P_2P_4) + P(P_4P_2) = **3 $\frac{1}{16}$**

Tabla 4.2: Espacio muestral generado por muestreo aleatorio con muestras de tamaño n = 2 con reposición, de una población de cuatro plantas de zapallo presentada en Tabla 4.1

Muestra	Plantas	Nro.de frutos	Media muestral	Muestra	Plantas	Nro.de frutos	Media muestral
1	P_1P_1	3; 3	3.0	9	P_3P_1	1; 3	2.0
2	P_1P_2	3; 2	2.5	10	P_3P_2	1; 2	1.5
3	P_1P_3	3; 1	2.0	11	P_3P_3	1; 1	1.0
4	P_1P_4	3; 4	3.5	12	P_3P_4	1; 4	2.5
5	$P_2 P_1$	2; 3	2.5	13	P_4P_1	4; 3	3.5
6	$P_2 P_2$	2; 2	2.0	14	P_4P_2	4; 2	3.0
7	$P_2 P_3$	2; 1	1.5	15	P_4P_3	4; 1	2.5
8	$P_2 P_4$	2; 4	3.0	16	P_4P_4	4; 4	4.0

Considérese ahora la tabla de frecuencias para la variable media muestral (Tabla 4.3). Obsérvese que la densidad no es uniforme (Figura 4.2) y que el valor más probable es 2.5, el cual corresponde a la esperanza de la distribución original de la variable número de frutos.

Tabla 4.3: Valores que asume la variable aleatoria "media muestral del número de frutos" en muestras de tamaño n=2 y sus densidades

Media Muestral	$P(\overline{X}=\overline{x})$
1	$1.\frac{1}{16} = 0.0625$
1.5	$2.\frac{1}{16} = 0.125$
2	$3.\frac{1}{16} = 0.1875$
2.5	$4.\frac{1}{16} = 0.25$
3	$3.\frac{1}{16} = 0.1875$
3.5	$2.\frac{1}{16} = 0.125$
4	$1.\frac{1}{16} = 0.0625$

Graficando la función de densidad de la media muestral para este ejemplo:

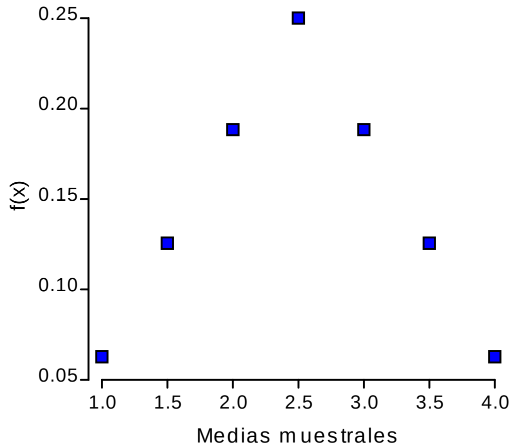


Figura 4.2: Función de densidad de la variable aleatoria media muestral del número de frutos obtenida por muestreo con reposición de tamaño n = 2 de una población de cuatro plantas de zapallo, presentada en tabla 4.1

Obsérvese que la esperanza de la distribución de las medias muestrales del ejemplo es igual a la esperanza de la distribución de la variable aleatoria original (número de frutos)

$$\mu_{\overline{X}} = 2.5 = \mu$$

Además la varianza de la distribución de las medias muestrales es igual a la varianza de la distribución de la variable estudiada, dividida por el tamaño muestral usado (en este caso n = 2).

$$\sigma^2_{\overline{X}} = \frac{\sigma^2}{n} = \frac{1.25}{2} = 0.625$$

Se usará la notación $\mu_{\overline{X}}$ y $\sigma^2_{\overline{X}}$ para representar a la esperanza y a la varianza de $\overline{X}$, respectivamente.

Definición 4.1: Error Estándar

La desviación estándar de las medias de muestras de tamaño n, recibe el nombre de **Error Estándar** y es definida como:

$$EE = \sqrt{\sigma^2_{\overline{X}}} = \sqrt{\sigma^2 / n}$$

Como se podrá observar, la varianza (y por ende el error estándar) de la variable

media muestral depende del tamaño de la muestra sobre la cual se calcula la media.

¿Cómo se pueden justificar los resultados anteriores?

En el ejemplo anterior, con una muestra de tamaño 2, se tienen dos variables aleatorias que se pueden designar con X_1 (número de frutos de la primer planta de la muestra) y X_2 (correspondiente de la segunda planta). Asumiendo que $E(X_1) = E(X_2) = \mu$ y que $V(X_1) = V(X_2) = \sigma^2$ y recordando que la esperanza y la varianza de combinaciones de variables aleatorias pueden expresarse como:

$$E(aX_1 + bX_2) = aE(X_1) + bE(X_2)$$

$$Var(aX_1 + bX_2) = a^2\ Var(X_1) + b^2\ Var(X_2)$$

Si *a* y *b* son constantes y X_1, X_2 variables aleatorias independientes:

Dado que $\overline{X} = \frac{1}{2} X_1 + \frac{1}{2} X_2$ entonces, usando las propiedades anteriores:

$$E(\overline{X}) = E(\tfrac{1}{2}X_1 + \tfrac{1}{2}X_2) = \tfrac{1}{2}E(X_1) + \tfrac{1}{2}E(X_2) = \tfrac{1}{2}\mu + \tfrac{1}{2}\mu = \mu$$

$$Var(\overline{X}) = Var\ (\tfrac{1}{2}X_1 + \tfrac{1}{2}X_2) = \left(\frac{1}{2}\right)^2 Var(X_1) + \left(\frac{1}{2}\right)^2 Var(X_2) =$$

$$= \frac{1}{4}\ \sigma^2 + \frac{1}{4}\ \sigma^2 = \frac{1}{4}\ (\sigma^2+\sigma^2) = \frac{1}{4}\ (2\sigma^2) = \boldsymbol{\sigma^2/2}$$

Se debe destacar el hecho de que la varianza de las medias muestrales es inversamente proporcional al tamaño de la muestra. Esto tiene un importante resultado práctico y es que **a través del tamaño muestral se puede controlar la variabilidad de la media resultante**. Consecuentemente, si la muestra es grande es menos probable que se obtenga una media muestral muy alejada de la esperanza de la distribución que se está muestreando, como puede observarse en la siguiente figura:

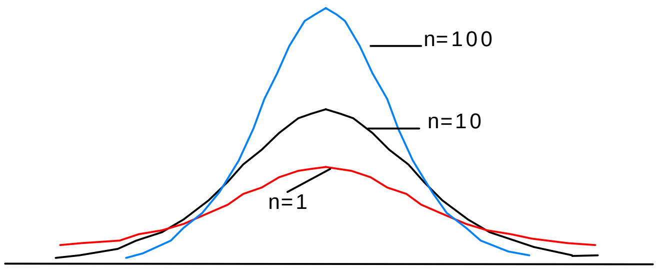


Figura 4.3: Funciones de densidad de la variable aleatoria media muestral basada en muestras de tamaño n = 1, n = 10 y n = 100 obtenidas de una población infinita

Un aspecto interesante de destacar en el Ejemplo 4.1 es la diferencia de la forma de la función de densidad de $\overline{X}$ presentada en la Figura 4.2, comparada con aquella de la variable original mostrada en la Figura 4.1. La Figura 4.2 muestra una densidad simétrica, centrada en μ, con forma triangular. Esta se asemeja más a la densidad normal que la densidad de la variable numero de frutos presentada en la Figura 4.1.

Si se hubieran utilizado muestras de mayor tamaño, se vería que la función de densidad se aproxima más aún a la gráfica de una densidad normal con idéntica esperanza y varianza inversamente proporcional al tamaño muestral. Este comportamiento no es casual sino la consecuencia de un importantísimo resultado que se resume en el siguiente teorema:

Teorema Central del Límite

Sea X una variable aleatoria con esperanza μ y varianza finita σ^2. Sea $\overline{X}$ la media muestral de una muestra aleatoria de tamaño *n* y Z la variable aleatoria definida como:

$$Z = \left(\frac{\overline{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \right)$$

entonces, la distribución de Z se aproxima a la distribución normal estándar cuando *n* se aproxima a infinito.

Note que el teorema no hace referencia a la distribución de X. Aunque X no se distribuya como una variable aleatoria normal, si tiene varianza finita, entonces para *"n"* suficientemente grande, la distribución de Z= $\left(\dfrac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}}\right)$ converge en distribución[4] a una N(0,1). Se dice entonces que Z posee una distribución *asintóticamente* normal.

El teorema central del límite provee un resultado muy importante ya que justifica la utilización de los métodos estadísticos que suponen normalidad en muchísimas situaciones prácticas.

Nota: Si la variable X se distribuye normal entonces Z = $\left(\dfrac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}}\right)$ tiene distribución exacta N(0,1) para cualquier tamaño muestral *"n"*.

Se ha visto que, dada una variable X con media μ y varianza σ^2, se puede derivar de manera aproximada o exacta la distribución de $\bar{X}$ haciendo uso del teorema central del límite. Luego, se puede calcular P($\bar{X} < \bar{x}$) o P($\bar{x}_1 < \bar{X} < \bar{x}_2$) como se mostró en el Capítulo 3 para variables aleatorias normales.

Si X ~ N (μ,σ^2) entonces para muestras de tamaño *"n"*:

$$P(\bar{X} < \bar{x}) = P\left(\frac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}} < \frac{\bar{x}-\mu}{\frac{\sigma}{\sqrt{n}}}\right) = P\left(Z < \frac{\bar{x}-\mu}{\frac{\sigma}{\sqrt{n}}}\right) \quad \text{donde } Z \sim N(0,1)$$

Distribución "T de Student"

La mayor dificultad en aplicar el resultado anterior es que, en la práctica, σ^2 es desconocida. Luego se podría estimar su valor a partir de una muestra, lo cual se logra sustituyendo en la fórmula anterior σ por el desvío estándar muestral.

El problema es que la sustitución de σ por S, modifica la variable aleatoria Z a la que

---

4 Cuando se dice que una variable con distribución $F_n(.)$ converge en distribución a una distribución G(.), cuando n tiende a infinito, se quiere indicar que $\forall\ \varepsilon > 0\ \exists\ n_0$ tal que $|F_n(x) - G(x)| < \varepsilon\ \forall\ x \in \Re$ si $n > n_0$

hace referencia el teorema central del límite y por tanto ya no se tiene una distribución normal para esta estandarización. La variable a aleatoria T = $\left(\frac{\overline{X}-\mu}{\frac{S}{\sqrt{n}}}\right)$ tiene una distribución conocida como T de Student con n-1 grados de libertad. Esta distribución es caracterizada por un único parámetro conocido como "grados de libertad" y que corresponde al número de observaciones que se utilizaron para calcular la desviación estándar muestral menos 1:

Luego, T= $\left(\frac{\overline{X}-\mu}{\frac{S}{\sqrt{n}}}\right)$ ~ T con (n-1) grados de libertad.

En consecuencia para calcular probabilidades del tipo P($\overline{X} < \overline{x}$) o P($\overline{x}_1 < \overline{X} < \overline{x}_2$), cuando no se conoce σ, se utiliza la "Tabla de Cuantiles de la Distribución T" (Tabla T), con los grados de libertad apropiados.

Cuando los grados de libertad de una distribución T son mayores que 30, la forma de la distribución, se aproxima a la de la distribución normal estándar. Es decir, ambas distribuciones están "suficientemente cerca" y por lo tanto utilizar en esos casos la Tabla T o la Tabla de cuantiles de la distribución normal estándar, produce resultados similares desde un punto de vista práctico. De hecho para grados de libertad infinitos la distribución T converge a la distribución normal.

Ejemplo 4.2

Considérese la variable *peso de 100 semillas* de una variedad de maíz. Para esta variable desconocemos la varianza aunque se puede suponer normalidad. El problema ahora es saber, para muestras de 5 paquetes de 100 semillas ¿cuál es la probabilidad de que la media muestral de los 5 paquetes sea menor de 38gr. si se supone que la esperanza de la distribución μ es 39gr.? Este ejemplo podría corresponder a la inquietud de un ente fiscalizador que desea saber cuál es el riesgo de que un lote bueno sea mal clasificado si se utiliza como criterio el peso promedio de 5 paquetes de 100 semillas.

Lo primero es tener una aproximación de σ^2, a través de la varianza muestral. Para eso se podrían tomar, por ejemplo, 12 bolsas de 100 semillas y pesarlas obteniendo los siguientes resultados:

37.4	38.0	40.2	37.9
39.1	38.5	41.0	37.7
38.2	39.4	39.9	40.1

En base a estos datos se obtiene S (el desvío estándar muestral) = $\sqrt{1.359}$ = 1.1658

Luego: $P(\overline{X} < 38 \text{ gr.}) = P\left(T < \dfrac{38-39}{\frac{1.1658}{\sqrt{5}}}\right) = P(T < -1.92) \cong 0.05$

donde T ~ T de Student con (12 - 1) grados de libertad. Los grados de libertad de la T se corresponden con el tamaño de la muestra con la que se calculó S.

Nota: Suponga que se quiere calcular $P[T \leq 4.3]$ donde T ~ T de Student con 2 grados de libertad. Tomando la fila de la Tabla de distribución T-Student (ver tabla en anexo), que corresponde a 2 grados de libertad se encuentra el valor 4.303 que corresponde a la columna encabezada por $t_{0.975}$. Esto indica que 4.3 es el cuantil 0.975 de la distribución T-Student de con 2 grados de libertad y en consecuencia $P[T \leq 4.3] = 0.975$. Si por el contrario la probabilidad requerida hubiera sido $P[T \leq -4.3]$ entonces se busca igualmente para t = 4.3 pero la lectura del cuantil se hará al pie de la columna debido al signo negativo del valor de la variable T. Luego, $P[T \leq -4.3] = 0.025$.

Distribución de la diferencia de dos medias muestrales

Tómense dos variables aleatorias X_1 y X_2 independientes que tienen distribución normal, tal que:

$$X_1 \sim N(\mu_1, \sigma_1^2) \text{ y } X_2 \sim N(\mu_2, \sigma_2^2)$$

Si el interés se centra en saber, por ejemplo, si las esperanzas de ambas distribuciones son idénticas, se podría definir una nueva variable aleatoria, como la diferencia entre X_1 y X_2 y estudiar el comportamiento de esta nueva variable a la que llamaremos *diferencia de dos variables aleatorias independientes* y se denota, en este ejemplo, por $(X_1 - X_2)$.

Se puede justificar que:

$E(X_1 - X_2) = E(X_1) - E(X_2) = \mu_{X_1} - \mu_{X_2}$

$\mathrm{Var}(X_1 - X_2) = \mathrm{Var}(X_1) + \mathrm{Var}(X_2) = \sigma^2_{X_1} + \sigma^2_{X_2}$

Si X_1 y X_2 son variables normales entonces la variable aleatoria diferencia $(X_1 - X_2)$ se distribuye también normalmente con esperanza $\mu_{X_1} - \mu_{X_2}$ y varianza $\sigma^2_{X_1} + \sigma^2_{X_2}$. En consecuencia para estandarizar la variable diferencia de dos variables aleatorias normales se tiene:

$$Z = \frac{(X_1 - X_2) - (\mu_1 - \mu_2)}{\sqrt{\sigma_1^2 + \sigma_2^2}} \qquad \text{con } Z \sim N(0,1).$$

Supóngase ahora, que se extraen muestras aleatorias de ambas distribuciones y, para cada una, se calcula la variable media muestral. Las distribuciones de estas medias muestrales son respectivamente:

$$\overline{X}_1 \sim N(\mu_1, \sigma_1^2/n_1) \quad \text{y} \quad \overline{X}_2 \sim N(\mu_2, \sigma_2^2/n_2)$$

Luego la variable aleatoria *diferencia de medias muestrales independientes* se distribuirá normalmente con:

$$E(\overline{X}_1 - \overline{X}_2) = E(\overline{X}_1) - E(\overline{X}_2) = \mu_1 - \mu_2$$

$$V(\overline{X}_1 - \overline{X}_2) = V(\overline{X}_1) + V(\overline{X}_2) = \sigma_1^2/n_1 + \sigma_2^2/n_2$$

Así la distribución de la diferencia de dos medias muestrales será:

$$(\overline{X}_1 - \overline{X}_2) \sim N\left(\mu_1 - \mu_2, \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}\right)$$

donde $\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}$ corresponde a la varianza de la diferencia de dos medias muestrales provenientes de dos distribuciones normales independientes. Luego, si se estandariza la diferencia de medias muestrales se tiene:

$$Z = \frac{\left(\overline{X}_1 - \overline{X}_2\right) - \left(\mu_1 - \mu_2\right)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \qquad \text{con } Z \sim N(0,1)$$

Si las distribuciones originales a partir de las cuales se obtuvieron $\overline{X}_1$ y $\overline{X}_2$ no son normales, se puede aplicar a esta diferencia las mismas propiedades que se deducen del teorema central del límite cuando n1 y n2 son ambas suficientemente grandes.

Nota: $\sqrt{\dfrac{\sigma_1^2}{n_1}+\dfrac{\sigma_2^2}{n_2}}$ es la desviación estándar de la variable diferencia de medias muestrales basadas en muestras de tamaño n_1 y n_2 También conocido como Error Estándar de la diferencia de dos medias muestrales.

Cuando no se conocen las varianzas distribucionales y se utilizan como sus aproximaciones a las varianzas muestrales, se deben reconocer dos situaciones:

a) las varianzas no se conocen, pero se saben iguales, en cuyo caso la desviación estándar de la diferencia de medias muestrales se calcula como:

$$S_{\overline{X}_1-\overline{X}_2}=\sqrt{S_p^2\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}, \quad \text{donde } S_p^2=\frac{(n_1-1)\,S_1^2+(n_2-1)\,S_2^2}{n_1+n_2-2}$$

Nota:

La expresión: $\sqrt{S_p^2}\cdot\sqrt{\dfrac{1}{n_1}+\dfrac{1}{n_2}}$ puede escribirse como: $S_p\sqrt{\dfrac{1}{n_1}+\dfrac{1}{n_2}}$ o $\sqrt{\left(S_p^2\cdot\dfrac{n_1+n_2}{n_1.n_2}\right)}$

b) las varianzas no se conocen pero se saben diferentes, en cuyo caso la desviación estándar de la diferencia de medias muestrales se calcula como:

$$S_{\overline{X}_1-\overline{X}_2}=\sqrt{\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}}$$

La "estandarización" que se obtiene utilizando una o otra expresión para el error estándar de la diferencia de medias, según sea el caso, es:

$$T=\frac{\left(\overline{X}_1-\overline{X}_2\right)-\left(\mu_1-\mu_2\right)}{S_{\overline{X}_1-\overline{X}_2}}$$

Esta expresión tiene distribución T-Student con (n_1 + n_2 - 2) grados de libertad en el caso "a" y distribución T-Student con $\nu = \dfrac{\left(\dfrac{S_1^2}{n_1} + \dfrac{S_2^2}{n_2}\right)^2}{\dfrac{\left(S_1^2/n_1\right)^2}{n_1+1} + \dfrac{\left(S_2^2/n_2\right)^2}{n_2+1}} - 2$ grados de libertad en el caso "b".

Ejemplo 4.3

Se tienen dos lotes de girasol y se toma de cada uno una muestra aleatoria simple de 10 paquetes de 100 semillas cada uno y luego se pesan. Los datos de peso de las bolsas de ambos lotes podrían ser las siguientes:

Lote 1	
43.3	46.7
55.6	42.8
46.8	47.9
56.3	43.7
45.7	34.6

Lote 2	
54.9	52.2
42.8	50.0
47.7	52.0
50.8	59.2
45.6	71.2

Lote	**n**	**Media**	**Varianza**
1	10	46.3	39.4
2	10	52.6	63.8

La diferencia de medias de la población 1 respecto a la 2 es 6.3 gramos. Luego se podría preguntar cuál es la probabilidad de que la diferencia de medias, basadas en muestras de tamaño 10 sea, por ejemplo, igual o mayor que la diferencia observada, si las esperanzas de las distribuciones de la variable peso de 100 semillas en ambos lotes fuera la misma (es decir μ_1-μ_2 = 0).

En términos de probabilidad, lo que se quiere averiguar es:

$$P\left(\left|\overline{X}_1-\overline{X}_2\right|\geq\left|\overline{x}_1-\overline{x}_2\right|\right),$$

La expresión anterior puede escribirse como:

$$P\left(\left(\overline{X}_1-\overline{X}_2\right)\leq-\left|\overline{x}_1-\overline{x}_2\right|\right)\ +\ P\left(\left(\overline{X}_1-\overline{X}_2\right)\geq\left|\overline{x}_1-\overline{x}_2\right|\right)$$

Suponiendo que $\mu_1 = \mu_2$, entonces:

$$P\left(\left(\overline{X}_1-\overline{X}_2\right)\geq\left|\overline{x}_1-\overline{x}_2\right|\right)\ =$$

$$P\left(\frac{\left(\overline{X}_1-\overline{X}_2\right)}{S_{\overline{X}_1-\overline{X}_2}}\leq\frac{-\left|\overline{x}_1-\overline{x}_2\right|}{S_{\overline{X}_1-\overline{X}_2}}\right)+P\left(\frac{\left(\overline{X}_1-\overline{X}_2\right)}{S_{\overline{X}_1-\overline{X}_2}}\geq\frac{\left|\overline{x}_1-\overline{x}_2\right|}{S_{\overline{X}_1-\overline{X}_2}}\right)$$

Asumiendo que las varianzas en ambas poblaciones son iguales, las probabilidades anteriores pueden reescribirse como:

$$P\left(T_{n_1+n_2-2}\leq\frac{-\left|\overline{x}_1-\overline{x}_2\right|}{S_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}\right)+P\left(T_{n_1+n_2-2}\geq\frac{\left|\overline{x}_1-\overline{x}_2\right|}{S_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}\right)$$

Haciendo los cálculos se tiene que la probabilidad buscada es aproximadamente 0.07. Esto quiere decir que la probabilidad de obtener una diferencia al menos tan grande como la observada es 0.07.

Distribución asociada al estadístico varianza muestral

Retomando el Ejemplo 4.1 que trataba con una población de 4 plantas de zapallos, donde la variable en estudio es la cantidad de zapallos en cada planta se vio que la varianza de la variable era $\sigma^2 = 1.25$.

Considérense, nuevamente, todas las muestras posibles de tamaño 2 obtenidas con reposición pero ahora en vez de calcular la media muestral, se calcula la varianza muestral para cada una de ellas. Los resultados se presentan en la siguiente tabla:

Tabla 4.4: Espacio muestral generado por muestreo aleatorio con muestras de tamaño 2 con reposición a partir de una población de cuatro plantas de zapallo, presentada en Tabla 4.1

Muestra	Plantas	Nº de frutos	Varianza	Muestra	Plantas	Nº de frutos	Varianza
1	P_1P_1	3-3	0.0	9	P_3P_1	1-3	2.0
2	P_1P_2	3-2	0.5	10	P_3P_2	1-2	0.5
3	P_1P_3	3-1	2.0	11	P_3P_3	1-1	0.0
4	P_1P_4	3-4	0.5	12	P_3P_4	1-4	4.5
5	P_2P_1	2-3	0.5	13	P_4P_1	4-3	0.5
6	P_2P_2	2-2	0.0	14	P_4P_2	4-2	2.0
7	P_2P_3	2-1	0.5	15	P_4P_3	4-1	4.5
8	P_2P_4	2-4	2.0	16	P_4P_4	4-4	0.0

En la siguiente tabla se presenta la distribución de la variable aleatoria *varianza muestral del número de frutos*.

Tabla 4.5: Valores que asume la variable aleatoria "varianza muestral del número de frutos" y sus densidades

Varianza muestral	$P(S^2 = s^2)$
0	$4.\frac{1}{16} = 0.25$
0.5	$6.\ \frac{1}{16} = 0.375$
2	$4.\ \frac{1}{16} = 0.25$
4.5	$2.\ \frac{1}{16} = 0.125$

Luego, la gráfica de la función de densidad es:

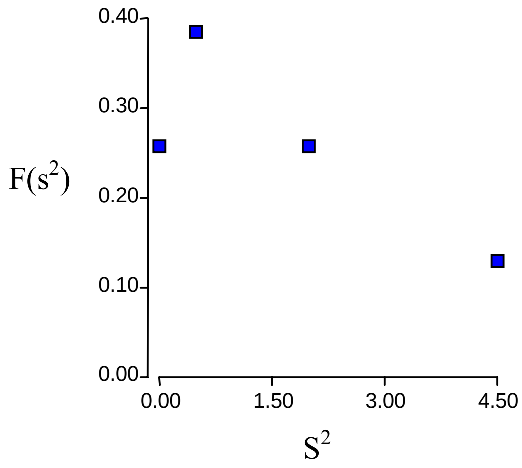


Figura 4.4: Distribución de frecuencias relativas de la variable aleatoria varianza muestral generada por muestreo con reposición de muestras de tamaño n=2 de una población de cuatro plantas de zapallo; presentada en Tabla 4.1

En la Tabla 4.4 se puede apreciar que la varianza muestral varía de muestra a muestra y en consecuencia la *varianza muestral*, es una variable aleatoria y como tal tiene una distribución asociada.

En la Figura 4.4 se ve que la distribución es **asimétrica**, con mayor concentración de valores a la izquierda de la media.

Para calcular probabilidades asociadas a varianzas muestrales se utiliza la distribución de la variable:

$$\frac{S^2\ (n-1)}{\sigma^2}$$

ya que se conoce que cuando S^2 es la varianza obtenida a partir de una muestra aleatoria de una distribución normal, la variable $\frac{S^2\ (n-1)}{\sigma^2}$, tiene distribución "Chi-cuadrado" con (n-1) grados de libertad.

Así, aplicando la propiedad de la esperanza de una variable aleatoria por una constante, E(cX) = cE(X), se tiene que:

$$E(S^2) = E\left(\frac{\sigma^2}{(n-1)}\frac{S^2\ (n-1)}{\sigma^2}\right) = E\left(c\frac{S^2\ (n-1)}{\sigma^2}\right)$$

donde $c = \dfrac{\sigma^2}{(n\text{-}1)}$.

Luego, como $\left(\dfrac{S^2\,(n-1)}{\sigma^2} \right)$ se distribuye como χ^2 con n –1 grados de libertad, y la esperanza de una variable aleatoria χ^2 es igual a sus grados de libertad,

$$E(S^2) = E(c\chi^2_{\,n-1}) = c(n\text{-}1) = \frac{\sigma^2}{(n-1)}(n-1) = \sigma^2$$

Obsérvese que la esperanza de la varianza muestral es igual a la varianza de la variable original. En el ejemplo: $E(S^2) = 1.25 = \sigma^2$. Luego, S^2 es un estimador insesgado de σ^2.

Ejemplo 4.4

Supóngase que la varianza máxima admisible para el peso de 100 semillas es 23 $gr.^2$ y que se obtiene desde una muestra de 10 paquetes de 100 semillas cada uno, una varianza muestral de $28gr^2$. ¿Es este resultado compatible con la especificación de la varianza máxima de $23gr^2$? Dicho desde un punto de vista estadístico se podría preguntar si la varianza muestral obtenida es un hecho frecuente o no, cuando la varianza de la distribución del peso de 100 semillas en la población que se está muestreando es a lo sumo $23gr^2$.

Luego se puede calcular,

$$P(S^2 \geq 28) = P(S^2\,(n-1)/\sigma^2 \geq 28\,(n-1)/\sigma^2) =$$
$$= 1 - P(S^2\,(n-1)/\sigma^2 \leq 28\,(n-1)/\sigma^2) =$$
$$= 1 - P(\chi^2 \leq 28\,(9)/23)) = 1 - P(\chi^2 \leq 10.96)$$

Buscando en la “Tabla de Cuantiles de la Distribución Chi-cuadrado” con 9 grados de libertad se encuentra que la probabilidad buscada es aproximadamente 0.75. Así 1-0.75 = 0.25

Luego, una de cada cuatro muestras de este tamaño tendrán varianzas iguales o mayores que 28, lo que para una población con varianza 23 es un resultado frecuente.

Ejercicios

Ejercicio 4.1

Al tirar un par de dados se obtienen realizaciones de dos variables aleatorias discretas independientes con valores posibles: {1,2,3,4,5,6}, cada uno de los cuales tiene probabilidad de 1/6.

a) ¿Cuál es la distribución de probabilidades de la variable *media del número de puntos en un par de dados*? Para responder, defina primero el conjunto de los resultados posibles de este experimento.

b) Graficar la densidad de la variable "*número de puntos en un dado"* y la densidad de la variable "*media del número de puntos en un par de dados".*

c) Comparar la forma de la densidad de la variable "*media del número de puntos en un par de dados"* con la forma de la densidade de la variable original ("*número de puntos en un dado")*.

Ejercicio 4.2

Si se especifica que la esperanza de la variable *cantidad de kilómetros recorridos por litros* de un vehículo es 12 y tiene una desviación estándar de 2.

¿Cuál es la probabilidad de que la media de una muestra de 10 recorridos sea menor o igual que 11 Km/lts si el vehículo funciona de acuerdo a las especificaciones?

Ejercicio 4.3

Si la distribución de la variable aleatoria producción de leche de un establecimiento lácteo (en cientos de litros) se aproxima a una distribución normal con media 70 y desvío estándar 8.

a) ¿Cuál es la probabilidad de que la media de una muestra de tamaño 5 exceda el valor 75?

b) ¿Cuál es la producción promedio sólo superada por un 5 % de las producciones promedio?

Ejercicio 4.4

Uso de la tabla de la Distribución "T" de Student

La tabla de la distribución T de Student del anexo contiene los cuantiles $t_{p,\nu}$ para algunos valores de p, con $p \in [0.55, 0.995]$ (encabezamiento de la tabla) y grados de libertad ν, con $\nu = 1, 2,...,50$.

Suponga que se quiere calcular la P(T ≤ 4.3) donde T es una variable aleatoria que tiene distribución T de Student con 2 grados de libertad. Se busca en el cuerpo de la tabla el valor 4.3 dentro de la fila que corresponde a $\nu = 2$, y en el encabezamiento de la columna se lee 0.975 que es la probabilidad buscada. El valor 4.3 es el cuantil 0.975 de la distribución T de Student con 2 grados de libertad.

Si por el contrario la probabilidad requerida hubiera sido P(T ≤ -4.3) entonces se procede de igual manera que en el párrafo anterior, pero la lectura de la probabilidad se hace en el pie de la columna. Luego P(T ≤ -4.3) = 0.025.

Obtener las siguientes probabilidades:

a) n = 50, P (T ≤ 2)

b) n = 50, P(T > 2)

c) n = 5, P(T ≤ -1.5)

d) ¿Cuál es el valor del cuantil 0.975 para una distribución T de Student con 5 grados de libertad? ¿Qué significa este valor?

e) ¿Cuál es el cuantil 0.30 para una distribución T de Student con 42 grados de libertad? ¿Qué significa este valor?

Ejercicio 4.5

Siguiendo con la situación planteada en el Ejercicio 4.3, responder las siguientes preguntas cuando no se conoce el valor de la desviación estándar de la distribución en estudio, y se dispone de la siguiente muestra para estimarla:

Muestra: 67.9 69.3 70.0 74.8 75.3 69.6 67.3 65.8 70.5

a) ¿Cuál es la probabilidad de que el estadístico $T = \left(\bar{X}_n - \mu\right)\Big/\sqrt{S^2/n}$ basado en la muestra de tamaño 9, exceda el valor 2?

b) ¿Cuál es la probabilidad de que el estadístico $T = \left(\bar{X}_n - \mu\right)\Big/\sqrt{S^2/n}$ basado en una muestra de tamaño 5, exceda el valor 2?

Ejercicio 4.6

Conocida la distribución de la media del número de puntos en un dado (Ejercicio 4.1), calcular la varianza muestral en cada uno de los pares de resultados posibles del experimento consistente en tirar un par de dados y registrar sus valores.

a) Construir la tabla de frecuencia para la variable varianza muestral y graficar su distribución.

b) ¿Cómo es la media de la distribución de varianzas muestrales respecto a la varianza de la variable original?

Ejercicio 4.7

Uso De la tabla de la Distribución Chi-cuadrado

En la tabla de distribución chi-cuadrado acumulada se pueden encontrar algunos cuantiles de la distribución para diferentes grados de libertad. Para calcular la probabilidad de que una variable distribuida como una chi-cuadrado con ν grados de libertad sea menor o igual a un cierto valor se procede de la siguiente forma:

Se busca en la tabla la fila que corresponde a los grados de libertad de la distribución y dentro de esa fila se localiza (de manera exacta o aproximada) el valor x. Luego se lee la probabilidad buscada mirando el encabezamiento de la columna correspondiente.

Por ejemplo, si X se distribuye como una χ^2 con 5 grados de libertad entonces:

$P (X \leq 3.99) = F (3.99) = 0.45$

Como ejercicio de uso de la tabla encontrar:

a) $P (X \leq 11)$ si X se distribuye como una χ^2 con 15 grados de libertad.

b) $P (S^2(n-1) /\sigma^2 \leq 4)$ si S^2 fue obtenido a partir de una muestra de tamaño 10.

Ejercicio 4.8

En un criadero de semillas se está probando una nueva variedad de maíz que saldrá a la venta si en una muestra de 50 parcelas experimentales el desvío estándar de su rendimiento no supera los 23 Kg/ha.

a) ¿Cuál es la probabilidad de que esto ocurra si la verdadera desviación estándar es 20?

b) ¿Cuál es el valor por debajo del cual está el 99% de los valores posibles de desviaciones estándar muestrales basadas en muestras de tamaño 30 si la verdadera desviación estándar es 20?

Ejercicio 4.9

La variable aleatoria *peso de latas de tomate* sigue una distribución normal. La desviación estándar de los pesos de latas de tomates en un lote de 10000 es igual a 1.4 grs.

Encontrar la probabilidad de que una muestra de 4 latas, tenga una desviación estándar que exceda 2.0 grs.

Ejercicio 4.10

Se sabe que la *longitud del fruto* de dos variedades (A y B) de tomate perita, sigue, en ambos casos, una distribución normal. Para la variedad A la media es $\mu = 7.3$ cm y la desviación estándar $\sigma = 0.4$ y para la especie B la media es de 6.0 cm y la desviación estándar 0.5 cm.

a) ¿Cuál es la distribución de la diferencia de medias muestrales de la longitud de frutos tomando $n_A = n_B = 5$?

b) ¿Cuál es la probabilidad de que la diferencia entre los promedios muestrales sea mayor o igual a 1.5 cm si $n_A = n_B = 10$?

c) ¿Qué proporción de la distribución de los promedios muestrales de la variedad B podría esperarse que estén comprendidos entre 5.5 y 6.5 cm con muestras de tamaño n=15?

5

Estimación de Parámetros

Introducción

Cuando se introdujo el concepto de Inferencia Estadística se indicó que una muestra de una población era útil para hacer inferencias acerca de la misma.

Dos importantes ramas de la Inferencia Estadística son la estimación de parámetros y la prueba de hipótesis. En este capítulo será tratado el problema de estimación y en el siguiente la prueba de hipótesis.

Los objetivos en este capítulo son: caracterizar las distribuciones de variables aleatorias a través de los parámetros media y varianza, estudiar el caso particular de la distribución normal, desarrollar la noción de estimación, presentar algunas propiedades de estimadores de parámetros y procedimientos para estimar parámetros.

Concepto de Estimación

En algunos casos se trata de estimar (aproximar numéricamente) la función de distribución de una variable aleatoria. Este es un objetivo ambicioso y puede requerir un esfuerzo muestral grande para lograr una buena estimación. En otros casos, se requiere que el investigador suponga la distribución de su variable y una vez establecida ésta, el problema es encontrar valores razonables para los parámetros que la caracterizan. Por ejemplo si la distribución supuesta es normal, los parámetros de interés podrían ser la esperanza y la varianza, ya que para especificar completamente la distribución es necesario conocer estos dos valores.
En el proceso de estimación de un parámetro hay dos enfoques que responden a diferentes necesidades: la *estimación puntual* y la *estimación por intervalo de confianza*.

Estimación Puntual

Cuando se aproxima un parámetro de una distribución a través de un valor decimos que se está haciendo es una estimación puntual. Supongamos que tenemos una muestra aleatoria (x_1, x_2, . . .,x_n) desde una distribución f(x;θ) y que deseamos usar esos valores para estimar el parámetro θ, el cual es desconocido. Luego, una función de x_1, x_2, . . .,x_n será usada para estimar θ.

Definición 5.1: Estimación y estimador puntual

Sea x_1, x_2, . . .,x_n una muestra aleatoria desde la distribución f(x; θ), la función $\hat{\theta}$ (x_1, x_2, . . .,x_n) es una estimación de θ . La función correspondiente de las variables aleatorias X_1, X_2, . . .,X_n, la cual es si misma una variable aleatoria, es un estimador puntual del parámetro θ .

Así, por ejemplo, la media muestral $\overline{X} = 1/n\sum_{i=1}^{n} X_i$ es una función de n variables aleatorias donde “n” es el tamaño de la muestra.

La calidad de la estimación obtenida depende de la adecuada elección del estimador puntual. Debido a que existe una gran variedad de estimadores posibles en cada situación particular es que necesitamos de criterios de selección.

Para seleccionar un buen estimador entre un conjunto de posibles estimadores, los estadísticos propuestos son estudiados teniendo en cuenta ciertas propiedades deseables.

Propiedades “clásicas” de los buenos estimadores

La elección de un estimador se realiza teniendo en cuenta, entre otros, los siguientes criterios:

a. Insesgamiento b. Consistencia c. Eficiencia

Insesgamiento

Definición 5.2: Insesgamiento

Un estimador $\hat{\theta}$ es un estimador **insesgado** para el parámetro θ si, para cualquier tamaño muestral, su esperanza es igual al parámetro que estima. Esto es, E($\hat{\theta}$) = θ , para todo valor de θ . El sesgo del estimador es definido como: Sesgo($\hat{\theta}$)=E($\hat{\theta}$ - θ).

Dicho de otra forma, si en promedio $\hat{\theta}$ = θ (la distribución de $\hat{\theta}$ esta centrada en θ o no existe una tendencia persistente a subestimar o sobreestimar θ), diremos que $\hat{\theta}$ es un estimador insesgado.

Esto se puede probar para la media muestral, de la siguiente manera: si se considera a la muestra de "n" observaciones como una colección de "n" variables aleatorias, todas idénticamente distribuidas con $E(X_i) = \mu \quad \forall_i$ luego,

$$E(\overline{X}) = E\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n}\sum_{i=1}^{n} E(X_i) = \frac{1}{n} n\mu = \mu$$

Es importante observar que la esperanza de la media muestral no depende de la distribución que se esté muestreando, sólo se pide que la distribución tenga esperanza.

Si se considera la varianza muestral se puede ver que este también es un estimador insesgado. Para $X \sim N(\mu,\sigma^2)$, $E(S^2) = \sigma^2$ ya que el estadístico $((n-1)\, S^2/\sigma^2) \sim \chi^2$ con (n-1) grados de libertad y en consecuencia se tiene que::

$$E(S^2) = E\left(\frac{\sigma^2}{(n-1)}\frac{S^2\ (n-1)}{\sigma^2}\right) = \frac{\sigma^2}{(n-1)} E\left(\frac{S^2\ (n-1)}{\sigma^2}\right) = \sigma^2$$

A diferencia de lo que ocurre con la media muestral, donde no se necesita suponer ninguna distribución para encontrar la esperanza, en el caso de la varianza muestral, se debe suponer normalidad para asegurar que $((n-1)\, S^2/\sigma^2)$ se distribuye como una variable "Chi-cuadrado" y desde allí obtener la esperanza como se mostró arriba.

La observación anterior es importante porque si se aplica la fórmula de varianza muestral para estimar la varianza de una distribución no normal, entonces no se puede asegurar que el estimador sea insesgado.

Consistencia

Definición 5.3: Consistencia

Sea $\hat{\theta}$ un estimador del parámetro θ, si la $P(|\hat{\theta} - \theta| > \varepsilon)$ tiende a 0, para $\forall$ ε>0, cuando el tamaño de la muestra tiende a ∞, se dice que $\hat{\theta}$ es un estimador **consistente** del parámetro θ.

En otras palabras, esto significa que a medida que aumenta el tamaño de muestra aumenta la proximidad de $\hat{\theta}$ respecto θ. Un ejemplo clásico de estimador consistente es la media muestral $\overline{X}$.

La consistencia es una característica esencial para cualquier estimador ya que implica que la calidad del resultado obtenido por la estimación refleja el esfuerzo muestral.

Eficiencia

Definición 5.4: Eficiencia

Sea $\hat{\theta}$ un estimador insesgado de θ, se dice que $\hat{\theta}$ es **eficiente** si tiene la mínima varianza posible.

Para comparar dos estimadores es útil el concepto de eficiencia relativa, que se obtiene desde la comparación de sus varianzas. Por ejemplo, si X es una variable aleatoria con distribución normal, entonces la media y la mediana muestral son estimadores insesgados del parámetro μ (la esperanza de la distribución), y además ambos estimadores son consistentes. Sin embargo, se puede comprobar que la varianza de la media muestral es menor que la varianza de la mediana por lo tanto la media es más eficiente que la mediana. Más aún, la media es **el estimador eficiente** en el sentido de que no existe ningún otro (bajo normalidad), que tenga menor varianza.

Sin embargo, cuando el supuesto de normalidad no se cumple, el estimador eficiente de la esperanza puede ser la mediana. Esto ocurre en distribuciones asimétricas o en distribuciones contaminadas (mezcla de distribuciones).

Estimación por Intervalo de confianza

Los estimadores puntuales son también variables aleatorias y, por lo tanto, no se puede esperar que en una realización cualesquiera den un valor idéntico al parámetro que estiman. Por ello, se desea que una estimación puntual esté acompañada de alguna medida del posible error de esa estimación. Esto puede hacerse indicando el error estándar del estimador o dando un intervalo que incluya al verdadero valor del parámetro con un cierto nivel de confianza.

Ejemplo 5.1

Si se quiere reportar el rendimiento de un cultivo, en vez de decir que la media del rendimiento se estima en 25 qq/ha se podría decir que, con una confianza del 95%, el rendimiento promedio para ese cultivo está comprendido entre 23.5 y 26.5 qq/ha.

El procedimiento que permite calcular los límites inferior y superior del intervalo antedicho se conoce como: **Estimación por Intervalo** y el intervalo obtenido: **Intervalo de Confianza**.

Procedimiento general para encontrar un intervalo de confianza para un parámetro.

El objetivo del procedimiento de estimación por intervalo es encontrar el intervalo cerrado [LI, LS] donde LI = Límite Inferior y LS = Límite Superior, tal que si el parámetro a estimar se simboliza por θ, entonces:

$$P(LI \leq \theta \leq LS) = 1- \alpha$$

Esta expresión se lee: "el intervalo de límites aleatorios LI y LS tiene probabilidad $(1-\alpha)$ de contener al parámetro θ", donde $(1-\alpha)$ denota la confianza de la estimación y se denomina **coeficiente de confianza**. Aunque la confianza se define como una cantidad que está entre 0 y 1, es frecuente expresarla como porcentaje, esto es: $(1-\alpha)\cdot 100$.

La especificación del coeficiente de confianza como $(1-\alpha)$ se hace por razones de consistencia con notación y conceptos que se introducirán posteriormente y en los que α tiene un significado particular.

Nota: Decir que un intervalo tiene confianza $(1 - \alpha)\cdot 100$ significa que: "si se utiliza el mismo procedimiento de construcción del intervalo para *m* muestras aleatorias independientes de idéntico tamaño *n*, entonces *m* $(1-\alpha)$ intervalos contendrán al verdadero valor del parámetro"**.**

Ejemplo 5.2

Si de una población con μ = 28, se toman 200 muestras independientes (m = 200) de tamaño "n" y se construyen para cada una un intervalo de confianza con coeficiente 0.90 (o del 90%), entonces se debe "esperar" que 180 de los 200 intervalos incluyan al valor 28.

Valores usuales de confianza son 0.95, 0.99 o 0.999. Estos niveles de confianza, aunque ampliamente aceptados, no constituyen una norma y pueden utilizarse otros.
Para poder construir estos intervalos se necesita:

a) Una función continua g(.,.) que relacione el parámetro θ y su estimador $\hat{\theta}$. Esto es $g(\theta, \hat{\theta})$.

b) Que $g(\theta,\hat{\theta})$ tenga una función de distribución F(.)[5] cuya especificación no dependa del parámetro θ.

Luego si $g(\theta,\hat{\theta})$ es la función que relaciona el parámetro y su estimador y F(.) su función de distribución, entonces:

$$P(q_1 \leq g\,(\theta,\hat{\theta}) \leq q_2) = 1-\alpha$$

implica que q_1 es el cuantil $(\alpha / 2)$ y q_2 el cuantil $(1 - \alpha / 2)$ de la distribución F(.). Una vez que se han establecido q_1 y q_2, los límites LI y LS surgen despejando θ a partir de $g(\theta,\hat{\theta})$.

A modo de ejemplo considérese este algoritmo aplicado a la estimación por intervalo de la esperanza y la varianza de una distribución normal.

Estimación de la esperanza de una variable aleatoria normal

Se deben distinguir dos casos dependiendo de si σ^2 es o no conocida.

Caso 1: Se conoce la varianza σ^2

Siguiendo los pasos descriptos en el procedimiento general se tiene:

- La función $g\,(\mu,\bar{X})$ para relacionar μ y su estimador $\bar{X}$ podría ser la siguiente:

$$g\,(\mu\,,\,\bar{X}) = (\bar{X} - \mu) \,/\, \sqrt{\sigma^2/n}$$

donde σ^2 es la varianza de la distribución y n el tamaño de la muestra a partir de la cual se hace la estimación.

- F (.) es, en este caso, N (0,1) ya que como se recordará (Capítulo 4):

$$(\bar{X} - \mu) \,/\, \sqrt{\sigma^2/n} \sim N\,(0,1)$$

Esta función es independiente del valor de μ, siempre y cuando μ sea la esperanza de la distribución.

Si se trabaja con una confianza del 95%, entonces $\alpha = 0.05$; $\alpha/2 = 0.025$;

$1 - \alpha/2 = 0.975$. Luego:

$$q_1 = Z_{(0.025)} = -1.96 \;\; y \;\; q_2 = Z_{(0.975)} = 1.96$$

---

5 No confundir esta función F(.) que indica una función de distribución genérica con la función F de Snedecor

que corresponden a los cuantiles 0.025 y 0.975 de una N (0,1).
Por lo tanto:

$$P(-1.96 \leq (\overline{X}-\mu) / \sqrt{\sigma^2/n} \leq 1.96) = 0.95, \text{ de donde:}$$

$$P(-1.96\sqrt{\sigma^2/n} \leq (\overline{X}-\mu) \leq 1.96\sqrt{\sigma^2/n}) = 0.95$$

Luego, restando $\overline{X}$:

$$P(-\overline{X} - 1.96\sqrt{\sigma^2/n} \leq -\mu \leq - \overline{X} + 1.96\sqrt{\sigma^2/n}) = 0.95$$

Multiplicando la expresión anterior por -1:

$$P(\overline{X} + 1.96\sqrt{\sigma^2/n} \geq \mu \geq \overline{X} - 1.96\sqrt{\sigma^2/n}) = 0.95$$

Reordenando:

$$P(\overline{X} - 1.96\sqrt{\sigma^2/n} \leq \mu \leq \overline{X} + 1.96\sqrt{\sigma^2/n}) = 0.95$$

Así:

$$LI = \overline{X} - 1.96\sqrt{\sigma^2/n} \quad y \quad LS = \overline{X} + 1.96\sqrt{\sigma^2/n}$$

Genéricamente se tiene:

$$\boxed{P(\overline{X} - Z_{1-\alpha/2} \cdot \sqrt{\sigma^2/n} \leq \mu \leq \overline{X} + Z_{1-\alpha/2} \cdot \sqrt{\sigma^2/n}) = (1 - \alpha)}$$

Caso 2: No se conoce la varianza σ^2

En el punto anterior se vio como encontrar el intervalo de confianza para μ cuando σ^2 era conocida. Sin embargo esta es una situación de interés solamente teórica ya que en general la varianza de la distribución es desconocida.

¿Cómo cambia el intervalo de confianza si se desconoce σ^2 ?

La función $g(\mu,\overline{X})$ que se utilizó en el punto anterior sufre una modificación que consiste en sustituir σ^2 por su estimador S^2, luego $g(\mu,\overline{X}) = (\overline{X} - \mu)/\sqrt{S^2/n}$.
Recuérdese (Capítulo 4) que esta sustitución del parámetro por su estimador produce cambios en la distribución dando como resultado que: $(\overline{X} - \mu) / \sqrt{S^2/n}$ **no** se distribuye N(0,1) sino como una $T_{(n-1)}$, donde (n-1) son los grados de libertad que caracterizan a esta distribución. Si se establece una confianza de $(1 - 0.05) \cdot 100 = 95\%$

y un tamaño muestral de por ejemplo n = 20, entonces, los cuantiles inferior y superior de una distribución T con (20 - 1) grados de libertad (g.l.) son: $q_1 = T_{\alpha/2} = -2.09$ y $q_2 = T_{1-\alpha/2} = 2.09$, respectivamente.

Por lo tanto:

$$P(-2.09\sqrt{S^2/n} \leq (\overline{X} - \mu) \leq 2.09\sqrt{S^2/n}) = 0.95$$

Luego, restando $\overline{X}$:

$$P(-\overline{X} - 2.09\sqrt{S^2/n} \leq -\mu \leq -\overline{X} + 2.09\sqrt{S^2/n}) = 0.95$$

Multiplicando por -1:

$$P(\overline{X} + 2.09\sqrt{S^2/n} \geq \mu \geq \overline{X} - 2.09\sqrt{S^2/n}) = 0.95$$

Reordenando:

$$P(\overline{X} - 2.09\sqrt{S^2/n} \leq \mu \leq \overline{X} + 2.09\sqrt{S^2/n}) = 0.95$$

Así:

$$\mathbf{LI} = \overline{X} - 2.09\sqrt{S^2/n} \quad \text{y} \quad \mathbf{LS} = \overline{X} + 2.09\sqrt{S^2/n}$$

Cálculo del tamaño muestral para obtener un intervalo de confianza para μ con una amplitud determinada

El problema que tratamos de resolver a continuación es establecer el tamaño de muestra necesario para obtener un intervalo de confianza para la esperanza de una distribución cuya amplitud sea menor o igual a una amplitud especificada por el investigador. En otras palabras, lo que se quiere es un método para obtener el tamaño muestral necesario para tener una estimación de la esperanza con la amplitud deseada.

Definición 5.5: Amplitud del intervalo de confianza.

Sean LI y LS los límites inferior y superior del intervalo de confianza para un parámetro θ. Luego la **amplitud (A) del intervalo de confianza** es A = LS - LI.

Nota: Los límites de un intervalo de confianza son aleatorios ya que se construyen en base a estadísticos muestrales.

Ejemplo 5.3

Si de una muestra aleatoria de tamaño 25 se obtiene: $\overline{X} = 12$ y S = 10 con el fin de calcular la amplitud, el intervalo de confianza (1 - α) para μ está dado por:

$$\mathbf{LS} = \bar{X} + T_{(n-1)\,;\,(1-\alpha/2)}\sqrt{S^2/n} \quad \text{y} \quad \mathbf{LI} = \bar{X} - T_{(n-1)\,;\,(1-\alpha/2)}\cdot\sqrt{S^2/n}$$

Entonces la amplitud es:

$$\mathbf{A = LS - LI} = \bar{X} + T_{(n-1)\,;\,(1-\alpha/2)}\sqrt{S^2/n} - \bar{X} + T_{(n-1)\,;\,(1-\alpha/2)}\sqrt{S^2/n}$$

y trabajando algebraicamente esta expresión y reemplazando con los valores propuestos queda:

$$\mathbf{A} = 2 \cdot T_{(24)\,;\,(0.975)} \cdot \sqrt{S^2/n} = 2 \cdot 2.064\sqrt{S^2/n} = 8.256$$

En este ejemplo, la amplitud es 8.256 unidades y se obtuvo con una muestra de tamaño 25. ¿Cuál debería ser el tamaño muestral para que la amplitud no supere las "c" unidades?, es decir, LS - LI ≤ c?

Este cálculo se realiza de manera sencilla haciendo:

$$A = 2 \cdot T_{(n-1),(1-\alpha/2)} \cdot \sqrt{S^2/n} \le c \quad \text{y despejando de allí "n"}$$

$$\text{Así, } n \ge \left(\frac{2\,.\,T_{(n-1);(1-\alpha/2)}\,.\,S}{c}\right)^2$$

Luego, para el Ejemplo 5.3, si c = 2 el tamaño muestral necesario será:

$$n \ge \left(\frac{2\,.\,2.064\,.\,10}{2}\right)^2 \cong 425$$

Nota: como el tamaño muestral fue calculado en base a una muestra preliminar y el coeficiente T (n-1),(1-α) depende de n, es recomendable hacer los cálculos con n = 425 (corrigiendo T (n-1),(1-α)) y luego recalculando "n". Así se tiene que T (424,0.95) corresponde al valor 1.96 y rehaciendo los cálculos se obtiene n = 384, ligeramente menor que el anteriormente calculado.

Si la amplitud quiere expresarse como una fracción "f" del valor medio, la expresión dada anteriormente se escribe como sigue:

$$n \ge \left(\frac{2\,.\,T_{(n-1);(1-\alpha/2)}\,.\,S}{\bar{X}.f}\right)^2$$

Nota: Los resultados presentados son aplicables para la construcción de intervalos de confianza aproximados para la esperanza de la distribución de variables aleatorias no normales siempre que sus distribuciones cumplan con los supuestos del teorema central del límite y "n" sea suficientemente grande.

Ejercicios

Ejercicio 5.1

Considerar la variable *rendimiento de maíz*, cuya distribución es normal con media μ y desviación estándar σ. Para estimar el rendimiento promedio del maíz bajo el efecto de un herbicida, se toma una muestra de tamaño 40 y se obtiene un promedio de 60 qq/ha. Se sabe por experiencias anteriores que la varianza poblacional σ^2 es 25 $(qq/ha)^2$.

a) Construir los intervalos de confianza del 95% y 99% para μ.

b) ¿Cómo cambia el intervalo anterior (95%) si el tamaño de la muestra fuese 100 y se obtiene el mismo promedio?

c) ¿Cómo se modifica el intervalo del 95% calculado en a) si la desviación estándar fuese de 7 qq/ha.?

Ejercicio 5.2

Una empresa dedicada a la comercialización de semillas desea estimar la altura promedio de un sorgo forrajero que ha desarrollado. Para ello toma una muestra de 50 plantas y se calcula la media de la altura, la que resulta ser 130 cm. Se sabe por experiencias anteriores que la desviación estándar es 22 cm.

Construir los intervalos de confianza para μ con una confianza del 95 % y 99 % respectivamente. Comparar ambos intervalos y concluir.

Ejercicio 5.3

Se quiere diseñar el tamaño de una muestra para estimar μ en una población normal con desviación estándar igual a 13.

a) ¿Cuál debería ser el tamaño mínimo de la muestra para asegurar una amplitud de 9 unidades para el intervalo de confianza al 95%?

b) ¿Qué sucede si la confianza cambia al 99%?

Ejercicio 5.4

Se desea establecer el contenido vitamínico de un alimento balanceado para pollos. Se toma una muestra de 49 bolsas y se encuentra que el contenido promedio de vitaminas por cada 100 grs. es de 12 mg. y que la desviación estándar es de 2 mg.

Encontrar el intervalo de confianza del 95% para el verdadero promedio del contenido de vitaminas.

Ejercicio 5.5

La distribución del rendimiento por ha. de una variedad de trigo en la zona de Leones tiene una media μ = 24.5 qq/ha. y una desviación estándar de 5 qq/ha. Se extraen 5 muestras de tamaño 100 cada uno, obteniendo las siguientes medias:

$$\overline{X}_1 = 24.1 \quad \overline{X}_2 = 25.5 \quad \overline{X}_3 = 23.0 \quad \overline{X}_4 = 24.0 \quad \text{y} \quad \overline{X}_5 = 25.9$$

a) Construir los intervalos de confianza del 95% para la media poblacional para cada uno de estos valores.

b) Considerar las cinco muestras como una única (de tamaño 500) y recalcular la media de esta muestra mayor ($\overline{\overline{X}}$) y el intervalo de confianza correspondiente.

c) ¿Se observa alguna diferencia entre la amplitud de los intervalos de las muestras individuales respecto de la amplitud del intervalo construido con la muestra mayor?

Ejercicio 5.6

El espárrago es una planta perenne cuyo cultivo comercial puede tener una duración de 15 años y su implantación es costosa. Dada la extensión del sistema radicular, la profundidad del suelo es fundamental, considerándose indispensable contar con un promedio mínimo de 80 cm de sustrato permeable. Se realizan 14 determinaciones de la profundidad del sustrato permeable (en cm) en puntos tomados al azar en dos campos (A y B). Los resultados fueron los siguientes:

A:	72	78	86	78	90	104	76	70	83	75	90	81	85	72
B:	78	82	68	68	74	81	85	73	75	89	100	91	82	75

A partir de los intervalos de confianza al 95% determinar si estos campos son aptos para el cultivo.

Ejercicio 5.7

Para estimar el rendimiento promedio del trigo en un departamento del sur cordobés se relevan los campos de distintos productores mediante un esquema de muestreo aleatorio simple. Se conoce por experiencias anteriores que σ es igual a 0.7 qq/ha y que el promedio histórico es 26 qq/ha.

a) ¿Qué número de campos se deben evaluar para estimar la media de rendimiento con una confianza del 95% si la amplitud del intervalo no debe ser mayor que el 2.5% del promedio histórico?

b) Si la varianza de la distribución aumenta (proponga σ = 1.4), ¿aumenta o disminuye el tamaño muestral necesario para mantener la misma amplitud? Justificar la respuesta.

6

Contraste de Hipótesis

Introducción

El hombre reconoce cotidianamente situaciones que le afectan, como la pérdida de cosechas, las enfermedades, las contingencias climáticas, etc. Tomar acciones para evitar o prevenir estos problemas requiere comprender cómo funciona el sistema que los origina. En el proceso de comprensión existe una etapa de idealización que se llama técnicamente *modelación*, que tiene por objeto identificar los elementos que son relevantes y plantear sus relaciones.

Si el modelo es correcto, en el sentido que representa bien el sistema bajo estudio, se tendrá una herramienta valiosa para planificar acciones en el mundo real.

¿Qué relación existe entre la construcción de estos modelos y la inferencia estadística?

Para que un modelo sea incorporado al patrimonio de la ciencia tiene antes que ser **validado,** es decir *mostrar que las predicciones que se deducen de él son aceptables*.

Lo usual es *realizar un experimento* u *observar el comportamiento del sistema* y comparar los resultados obtenidos en estos estudios con los que se deducen del modelo. Si no hay **diferencias significativas** entre lo **observado** y lo **esperado**, entonces se dirá que el modelo es correcto para esa situación (o desde un punto de vista más estricto: que el modelo es "provisoriamente" aceptable).

El problema es definir qué se entiende por **diferencia significativa**. No es simple establecer un criterio para decir si la discrepancia entre lo que se observa y lo que se espera es grande o pequeña.

Por ejemplo, si un modelo de precipitaciones predice que en los primeros 10 días del mes de enero lloverá 60 mm en una localidad de la Provincia de Córdoba y en cambio se registran 40 mm, para un detractor del modelo la diferencia será significativa mientras que para otros no lo será. ¿Cómo ser imparcial en este juicio?

En primer lugar se deberá discutir si es razonable aceptar que *el milimetraje de lluvia caída en los 10 primeros días de enero* se puede tratar como una variable aleatoria. Si se concluyera afirmativamente, entonces, basándose en el modelo propuesto para las precipitaciones, se podría derivar su distribución y a partir de ella asignar una probabilidad al evento: "obtener un milimetraje de más de 20 milímetros por debajo o por encima de la esperanza de la distribución de lluvias" (el milimetraje predicho por el modelo, en este caso, es de 60 mm). Con esta **medida** de probabilidad se podrá tomar una decisión que es reproducible por cualquier investigador. Para el caso, considérese un ejemplo extremo: suponga que al calcular esta probabilidad se obtiene que el evento tiene una chance de ocurrir 1 de cada 1.000.000 de veces (uno en un millón). Esto quiere decir que de cada un millón de períodos que van del 1 al 10 de enero en la localidad citada, sólo uno tendrá un milimetraje que discrepa en 20 o más milímetros de lo esperado bajo el modelo. Luego, con la evidencia observada se pueden sacar dos conclusiones: 1) que se tuvo muy mala suerte (justo se observó el período que ocurre una vez cada millón de años), o 2) que el modelo es incorrecto. Lo usual, en estos casos, es aceptar la segunda alternativa.

Otro hubiera sido el caso si la probabilidad del evento mencionado fuera 0.40, esto es 40 de cada 100 años ocurren discrepancias iguales o mayores que la observada. Aquí, la evidencia muestral no tiene peso suficiente para que se rechace el modelo ya que el evento observado es un evento frecuente.

La idea es entonces: dado un modelo no validado - que se llama **hipótesis científica** - se debe seguir algún procedimiento para deducir alguna consecuencia, cuya verificación o falta de verificación, sirva para establecer la veracidad de la hipótesis científica. Si la/s consecuencia/s de la hipótesis científica se pueden visualizar como propiedades estadísticas de una **variable aleatoria,** será factible utilizar herramientas estadísticas para tomar una decisión sobre la veracidad del modelo. Para ello se debe expresar la hipótesis científica como una **hipótesis estadística**. Estas hipótesis consisten en una afirmación sobre uno o más parámetros de la distribución de la variable aleatoria en cuestión, como sería por ejemplo, para la variable *milimetraje de lluvia* indicar que la esperanza de la distribución $\mu = 60$ mm.

Es obvio que la hipótesis estadística debe ser equivalente a la hipótesis científica postulada, de lo contrario, aceptar o rechazar la hipótesis estadística no implicará necesariamente lo propio para la hipótesis científica.

Conceptualmente la prueba estadística o **prueba de hipótesis** es sencilla: se examina un conjunto de datos muestrales y a partir de ellos se calcula un estadístico cuya distribución depende de la hipótesis planteada. Sobre la base de la distribución especificada para el estadístico y de su valor observado en la muestra, se decide el rechazo o no de la hipótesis estadística y en consecuencia de la hipótesis científica.

Aunque las hipótesis científicas pueden dar lugar a hipótesis estadísticas que involucran a más de un parámetro de la distribución de una o más variables aleatorias, la discusión que sigue se limitará, por razones de simplicidad, al caso de pruebas de hipótesis acerca de un parámetro de la distribución de una variable aleatoria normal.

Los objetivos de este capítulo son establecer relaciones entre el Contraste de Hipótesis y el Método Científico, analizar las etapas fundamentales de la Prueba Estadística de Hipótesis, conceptualizar los distintos tipos de errores, y establecer relaciones con la Estimación de Parámetros.

Procedimiento de la Prueba de Hipótesis

A fin de dar una idea general de la metodología de la prueba estadística de hipótesis, y aunque se incluyen conceptos que se definen posteriormente, a continuación se enumeran los pasos a seguir en la prueba de una hipótesis estadística:

a) Plantear las hipótesis nula y alternativa.

b) Planificar el experimento o el esquema muestral conducente a obtener datos que permitan la validación o no de la hipótesis sometida a prueba.

c) Seleccionar (o construir) un estadístico cuya distribución quede completamente especificada bajo la hipótesis nula[6].

d) Establecer el nivel de significación de la prueba.

e) Establecer los eventos que conducen al rechazo y no rechazo de la hipótesis nula mediante la definición de regiones de rechazo y de no rechazo (aceptación).

f) Realizar el ensayo o muestreo *"ad hoc",* definido en el punto b para obtener las observaciones con las que se realizará la prueba.

g) Calcular el valor del estadístico postulado y determinar si está dentro o fuera de la región de rechazo. En el primer caso se dice que se rechaza la hipótesis nula y en el segundo que no.

El orden en que se presentan los pasos anteriores es una secuencia formal que no siempre se respeta en la práctica de la investigación. Usualmente se tiene una hipótesis científica y se planifica una experiencia para probarla y una vez obtenidos los datos se trata de formalizar una hipótesis estadística. Debe advertirse que aunque en la práctica es usual este proceder, decididamente no es recomendable ya que la elección del estadístico y su distribución dependen de la forma en que se planifica el experimento

6 Con la expresión “bajo hipótesis nula” se indica “suponiendo que lo que especifica la hipótesis nula es cierto”

(o el muestreo) y de la naturaleza de la hipótesis estadística formulada. Si esto no se ha tenido en cuenta a la hora de planificar la experiencia, puede ocurrir (y de hecho ocurre con mucha frecuencia) que los datos obtenidos sean de escaso o nulo valor para realizar una prueba estadística.

A continuación se definen y discuten cada uno los pasos presentados.

Plantear las hipótesis nula y alternativa

Para poder construir una prueba estadística se debe especificar una hipótesis que se supone, provisoriamente como verdadera, llamada **hipótesis nula** y es simbolizada con H_0. Esta hipótesis especifica los valores de uno o varios parámetros de la distribución de la variable aleatoria observada en el experimento. Cuando la hipótesis nula se somete a prueba, el resultado es su aceptación o rechazo. En este último caso se aceptará una hipótesis especificada de antemano que se llama **hipótesis alternativa**, que se simboliza por H_1 y que propone como posibles valores del o los parámetros en cuestión al conjunto de valores complementarios al postulado bajo H_0.

Planificar el experimento o el esquema muestral

La forma en que se recolectan los datos o se diseña el experimento es motivo de tratamiento particular por las técnicas de muestreo y el diseño de experimentos. A modo de introducción se puede decir que el objetivo de este paso es definir la forma en que los datos serán obtenidos, incluyendo el número total de observaciones en la muestra (o el número de repeticiones del experimento).

Selección de un estadístico para la prueba e identificación de su distribución bajo H_0

El estadístico de la prueba es una función de la muestra. Se necesita una función W de la muestra cuya distribución sea conocida y quede completamente especificada bajo H_0, es decir que se puede calcular $P(W \leq w)$. La función W, a través de su distribución, servirá para asignar probabilidades a los eventos que conducen a aceptar o rechazar la hipótesis nula postulada. El evento que induce al rechazo se conoce como *región o zona de rechazo de H_0*, en tanto que el evento que conduce al *no rechazo* se llama *región o zona de aceptación de H_0*.

Nota: entre todos los estadísticos posibles para una prueba de hipótesis se recomienda elegir aquel que maximiza la potencia de la prueba (ver definición de potencia más adelante).

Establecer el nivel de significación de la prueba

Definición 6.1: Nivel de significación

El **nivel de significación** se define como la máxima probabilidad de rechazar H_0 cuando ésta es verdadera. Será denotado por la letra griega α.

El nivel de significación representa la máxima probabilidad de equivocarse en el sentido de ***concluir que H_0 es falsa cuando en realidad no lo es***. Este error, llamado **Error de Tipo I**, será considerado detenidamente en la próxima sección.

Una vez que se han establecido H_0 y H_1 debe fijarse el **nivel de significación.** En general se fija en 0.05 (5%) o en 0.01 (1%), que son niveles usualmente aceptados, aunque no hay razón alguna para no seleccionar algún otro.

Es importante indicar que la probabilidad de cometer el error de tipo I se establece antes de la realización de la prueba estadística. Esta observación tiene el objetivo de que el investigador evalúe cuál es la tasa de error de tipo I que está dispuesto a tolerar en base a criterios independientes de los resultados muestrales o experimentales.

Establecer los eventos que conducen al rechazo y no rechazo de la Hipótesis Nula

Una vez fijados el estadístico de la prueba, su distribución y el nivel de significación, el próximo paso consiste en establecer las *regiones de no rechazo y de rechazo de H_0.*

Definición 6.2: Región o zona de rechazo

La **región de rechazo** de H_0 es uno o más intervalos de la recta real que describen al evento que conduce al rechazo de H_0 y cuya probabilidad, cuando H_0 es verdadera, es α.

Definición 6.3: Región o zona de no rechazo

La **región de no rechazo** de H_0 es un intervalo de la recta real que describe al evento que conduce al no rechazo de H_0 con probabilidad $1-\alpha$,cuando H_0 es cierta.

La zona de rechazo puede estar a la izquierda o a la derecha de la distribución del estadístico bajo H_0, y en estos casos se dice que la prueba es **unilateral izquierda** o **derecha** respectivamente. Cuando la zona de rechazo está repartida a izquierda y derecha se dice que la prueba es **bilateral**.

La condición bilateral o unilateral de la prueba de hipótesis depende de la hipótesis alternativa. El "tamaño" de la región de rechazo esta determinado por el nivel de significación de la prueba.

Así, si la hipótesis nula es $\mu = \mu_0$ y su alternativa es de la forma $\mu \neq \mu_0$ se está en presencia de una **prueba bilateral** y la zona de rechazo estará ubicada en las dos colas (izquierda y derecha) de la distribución del estadístico de la prueba. Si el nivel de significación fuera del 5% ($\alpha = 0.05$), las "porciones" derecha e izquierda de la zona de rechazo tendrán asociadas una probabilidad de 0.025 cada una[7].

En contraposición, si la hipótesis alternativa es $\mu > \mu_0$ o $\mu < \mu_0$ la prueba es **unilateral derecha** o **izquierda** respectivamente**,** y la zona de rechazo de H_0 estará ubicada en la cola derecha o izquierda de la distribución del estadístico de la prueba y la probabilidad asociada a la región será "α" en lugar de "$\alpha/2$" como en las pruebas bilaterales.

Ejemplos de hipótesis:

Bilateral	**Unilateral derecha**	**Unilateral izquierda**
H_0: μ = 20 qq/ha	H_0: $\mu \leq$ 20 qq/ha	H_0: $\mu \geq$ 20 qq/ha
H_1: $\mu \neq$ 20qq/ha	H_1: $\mu >$ 20 qq/ha	H_1: $\mu <$ 20 qq/ha

Definición 6.4: Puntos críticos

Los valores de la recta real que separan la zona de no rechazo de la de rechazo se denominan **puntos críticos**.

Las Figuras 6.1-6.3 ejemplifican los distintos casos de pruebas de hipótesis, donde se señalan las zonas de no rechazo y rechazo, los puntos críticos y las probabilidades asociadas al rechazo de H_0. Como se observa, el estadístico utilizado en estos casos tiene esperanza 0 (cero) y una distribución simétrica (como podría ser la distribución normal o la T de Student).

---

[7] La división de la zona de rechazo en dos zonas de idéntico tamaño en términos de probabilidad puede parecer arbitraria, pero es la que garantiza la máxima potencia.

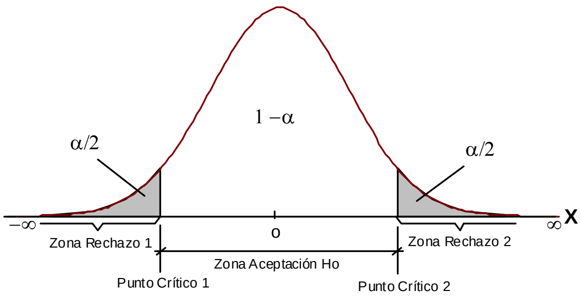


Figura 6.1: Representación de la distribución del estadístico bajo H_0 en una prueba bilateral

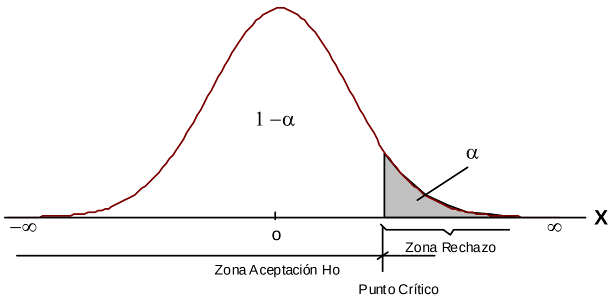


Figura 6.2: Representación de la distribución del estadístico bajo H_0 en una prueba unilateral derecha

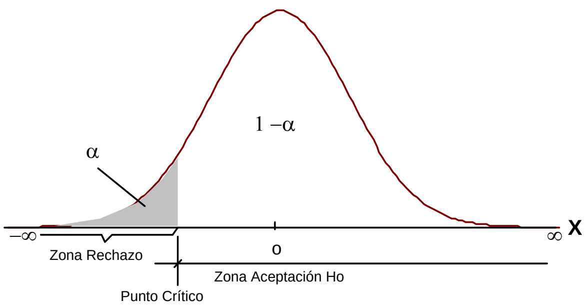


Figura 6.3: Representación de la distribución del estadístico bajo H_0 en una prueba unilateral izquierda

Realizar un ensayo o muestreo "*ad Hoc*"

Esta etapa tiene por objeto obtener datos experimentales que permitan evaluar el estadístico propuesto para la prueba, de acuerdo a la planificación realizada previamente.

Calcular el valor del estadístico y determinar si está dentro o fuera de la región de rechazo

Con los datos obtenidos en el paso anterior se calculará W, cuya distribución bajo la hipótesis nula es conocida y para la cual se han fijado las regiones de no rechazo y rechazo.

Si el valor calculado de W pertenece a la región de rechazo se concluye que la hipótesis nula debe desecharse. En caso contrario se concluye que no hay evidencia suficiente (o como se verá más adelante, quizás suficiente potencia), para rechazarla.

Ejemplo 6.1

Se desea probar si una nueva variedad de soja lograda por un proceso de mejoramiento genético supera la base de 20 qq/ha.

De acuerdo a los pasos enunciados anteriormente se tiene:

Paso 1: Planteo de la hipótesis estadística

H_0: $\mu \le \mu_0$ (20 qq/ha)

H_1: $\mu > \mu_0$ (20 qq/ha).

Paso 2: Para probar la hipótesis se planifica una experiencia que consiste en repetir el cultivo de la nueva variedad de soja en 30 parcelas de ¼ de hectárea cada una, y registrar sus rendimientos a cosecha.

Paso 3: Elección de la función de la muestra W

Se tomará $W(X_1,...,X_n) = \dfrac{(\bar{X}-\mu_0)}{\sqrt{\dfrac{S^2}{n}}}$

Este estadístico, cuando $\mu = \mu_0$ y bajo el supuesto de normalidad para la variable en estudio, se distribuye como una T de Student con n-1 grados de libertad.

Paso 4: Se fija el nivel de significación α, por ejemplo, en 0.05.

Paso 5: Para el establecimiento de las zonas de aceptación y de rechazo, es necesario establecer el punto crítico (PC). El PC para W, que en este ejemplo es una variable T de Student, es el cuantil 0.95 de la distribución T con (n-1) grados de libertad y se denota como $T_{(n-1);0.95}$.

Si de acuerdo a la planificación del experimento hay 30 repeticiones, en la tabla t de Student se obtiene PC = 1.699, por lo tanto la zona de aceptación de H_0 (ZA) y la de rechazo de la H_0 (ZR) serán:

$$ZA = (-\infty\ , 1.699) \text{ y } ZR = [1.699, \infty)$$

<u>Paso 6:</u> De acuerdo a la planificación del experimento, referida en el Paso 2, al cabo de la cosecha se obtienen los siguientes resultados: $\overline{X}$ = 25 qq/ha y S = 4 qq/ha.

<u>Paso 7</u>: Se calcula W y se observa a cuál de los intervalos definidos en el Paso 5 pertenece. En función de ello se acepta o no la hipótesis, así:

$$W(x_1 \ldots. x_n) = \frac{25 - 20}{\frac{4}{\sqrt{30}}} \approx 6.847$$

Como $W \in ZR$ se concluye que se rechaza la hipótesis nula, por lo tanto la nueva variedad supera en promedio el rendimiento de 20 qq/ha.

Es importante observar que si se construye un intervalo de confianza del $(1-\alpha)\cdot 100$ unilateral izquierdo para μ, el límite inferior sería mayor que 20 qq/ha, lo cual es consistente con lo encontrado en la prueba de hipótesis. De hecho, toda prueba tiene asociada un intervalo de confianza y viceversa.

Errores

En la prueba de una hipótesis estadística pueden ocurrir dos errores: el error de tipo I y el error de tipo II. La posible ocurrencia de uno u otro error depende de la condición de verdadera o falsa de la hipótesis nula y de la decisión, basada en la muestra, de aceptarla o rechazarla.

Si se rechaza H_0, el **Error de Tipo I** se comete cuando H_0 es verdadera. Si por el contrario se acepta H_0 siendo ésta falsa, entonces se cometerá el **Error de Tipo II**. Es importante advertir que estos errores constituyen eventos de espacios muestrales diferentes definidos por la condición de verdadera o falsa de H_0. En consecuencia, si se asignan a ellos probabilidades de ocurrencia α y β para el error tipo I y II respectivamente, **será incorrecto decir que la probabilidad de error total en una prueba de hipótesis es la suma $\alpha + \beta$.**

Con la finalidad de dar una interpretación de α, considérese el siguiente ejemplo.

Ejemplo 6.2

Una muestra de 25 observaciones procede de una distribución normal con media μ=50 y desviación estándar σ = 10, por lo tanto la distribución muestral de $\overline{X}$ es normal con media μ = 50 y desviación estándar $\sigma / \sqrt{n}$ = 10/5 = 2.

Si el interés se centra en probar la hipótesis nula μ = 50 vs. $\mu \neq 50$, luego, se toma como estadístico de la prueba a $W(X_1,...,X_n) = \dfrac{(\bar{X}-\mu)}{\sqrt{\dfrac{\sigma 2}{n}}}$ cuya distribución es N(0,1) cuando H_0 es cierta. Para encontrar los valores de W que conducen al rechazo de H_0, se obtienen los puntos críticos de la tabla de distribución normal teniendo en cuenta que la hipótesis alternativa implica una prueba bilateral y tomando un nivel de significación del 5%. Luego:

$$PC_1 = Z_{(\alpha/2)} = -1.960 \;\; y \;\; PC_2 = Z_{(1-\alpha/2)} = 1.960$$

Así, la regla de decisión es: se rechaza μ = 50 si W es menor que -1.960 o si W es mayor que 1.960 y no se rechaza μ = 50 si W está entre -1.960 y 1.960.

Nótese que la tabla de los cuantiles de la distribución normal estándar muestra que el 2,5% del área por debajo de la curva se corresponde con valores menores de -1.960 y otro porcentaje similar con valores mayores que 1.960.

En síntesis:

Cuando H_0 es cierta el 5% de las muestras tendrán un valor de W menor que -1.960 o mayor que 1.960, y para estas muestras la anterior regla de decisión conducirá al error de rechazar que μ es 50.

Por otra parte, el 95% de las muestras presentarán valores de W entre -1.960 y 1.960 y en estos casos se decidirá correctamente que μ = 50.

Cálculo de la probabilidad de cometer Error de Tipo II (β)

Prosiguiendo con el ejemplo, supóngase que la hipótesis nula no es verdadera y que μ = 52. Lo que interesaría saber, en este caso, es cuál es la probabilidad de aceptar H_0 cuando es falsa. Para conocer esta probabilidad se debe encontrar la probabilidad de la región de aceptación cuando μ = 52. En el ejemplo, $P(-1.960 \leq W \leq 1.960 \mid \mu = 52)$.

Para hallar esta probabilidad se debe conocer la distribución de W cuando $\mu = 52$. De manera general $W \sim N\left(\dfrac{\mu-\mu_0}{\sqrt{\dfrac{\sigma^2}{n}}}, 1\right)$ donde μ_0 es la esperanza bajo H_0 y μ la verdadera esperanza de la distribución. Luego, utilizando la expresión anterior se tiene que $W \sim N(1,1)$ cuando $\mu = 52$. En consecuencia β, la probabilidad de Error de Tipo II, que corresponde a la probabilidad de la región de aceptación bajo la hipótesis alternativa, está dada en este caso por:

$\beta = P(-1.960 \leq W \leq 1.960 \mid \mu = 52) = P((-1.960 - 1) / 1 \leq Z \leq (1.960 - 1) / 1))$

donde $Z \sim N(0,1)$

Por lo tanto, si la hipótesis nula no es verdadera y $\mu = 52$, entonces $\beta = 0.83$. Si se toman sucesivas muestra de tamaño 25 de una población con $\mu = 52$ (en lugar de 50) y $\sigma = 10$, la regla de decisión que se ha usado con $\alpha = 0.05$ conducirá a aceptar incorrectamente la hipótesis nula 83 de cada 100 veces; esto se ilustra en la Figura 6.4.

Para la prueba de hipótesis acerca de la esperanza de una distribución normal y para un nivel de significación α dado, β se calcula según las siguientes expresiones:

$\beta = P\left(Z \leq Z_{1-\alpha} - \dfrac{\mu-\mu_0}{\sqrt{\dfrac{\sigma^2}{n}}}\right)$ si la prueba es unilateral derecha y

$\beta = P\left(Z \geq Z_{\alpha} - \dfrac{\mu-\mu_0}{\sqrt{\dfrac{\sigma^2}{n}}}\right)$ si la prueba es unilateral izquierda.

Si la prueba es bilateral entonces:

$$\beta = P\left(Z_{\alpha/2} - \frac{\mu - \mu_0}{\sqrt{\frac{\sigma^2}{n}}} \leq Z \leq Z_{1-\alpha/2} - \frac{\mu - \mu_0}{\sqrt{\frac{\sigma^2}{n}}} \right) \text{donde Z} \sim \text{N (0,1)}$$

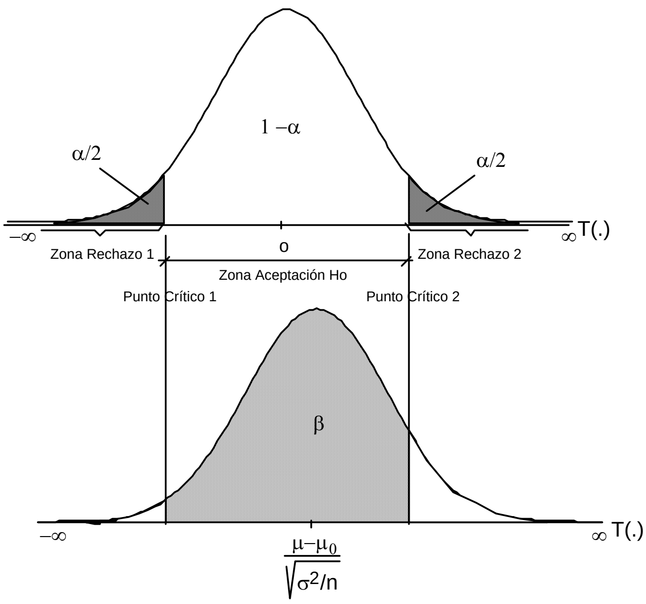


Figura 6.4: Áreas asociadas a los errores

Por lo general el investigador no determina la probabilidad de cometer el error de tipo II, aunque el ejemplo anterior debe advertir sobre la importancia de tenerla en cuenta. Usualmente se prueba H_0 fijando solamente α pero, como se verá más adelante, el manejo del tamaño muestral o el número de repeticiones de un experimento, es el elemento a modificar para controlar β. La importancia relativa de los errores depende de los costos inherentes a cada tipo de error y estos costos deberían servir como pautas para fijar las probabilidades de cometerlos.

Las posibles decisiones y sus errores, concernientes a la prueba de hipótesis, se resumen en la siguiente tabla:

Tabla 6.1: Probabilidades asociadas a las distintas decisiones en la prueba de hipótesis.

Decisión	Error	Probabilidad
Si H_0 Cierta y:		
Se rechaza H_0	Tipo I	α
No se rechaza H_0	Nulo	$1-\alpha$
Si H_0 Falsa y:		
Se rechaza H_0	Nulo	$1-\beta$
No se rechaza H_0	Tipo II	β

Efectos de las variaciones de la región de rechazo sobre β

La afirmación que β aumenta según disminuye α es verdadera para "n" fijo. El investigador que hace el experimento quizás desee variar el nivel de significación de la prueba para obtener la correspondiente variación en β. Un valor chico de α es deseable, pero tomarlo demasiado pequeño puede hacer β tan grande que se tenga muy poca chance de reconocer si la hipótesis nula es falsa.

Efecto de las variaciones del tamaño de la muestra sobre β

Manteniendo constante el nivel de significación, la región de aceptación es más pequeña para tamaños de muestras mayores con la consiguiente disminución de β. En la Tabla 6.2 se dan los valores de β para la alternativa $\mu = 52$, con hipótesis nula $\mu=50$, mostrando numéricamente el efecto de cambiar α y *n* sobre la probabilidad del Error de Tipo II.

Tabla 6.2: Valores de β para n y α dados.

n	$\alpha = 0.05$	$\alpha = 0.01$
1	0.945	0.988
3	0.936	0.985
5	0.927	0.980
10	0.903	0.973
25	0.830	0.942
50	0.707	0.877
100	0.484	0.717
400	0.021	0.077
1000	0.00001	0.0001

Potencia de una prueba de hipótesis

Definición 6.5: Potencia de una prueba

Se define como **potencia** a la probabilidad de rechazar la Hipótesis Nula cuando ésta es falsa. La potencia se denota como π.

Esta probabilidad representa la chance de concluir que H_0 es falsa cuando efectivamente lo es. La potencia se calcula como $\pi = 1 - \beta$, donde β es la probabilidad de cometer el Error de Tipo II. Cuanto mayor es la potencia mejor es la prueba. La potencia es función de varios factores: a) el nivel de significación elegido, b) la varianza de la variable aleatoria y c) el tamaño de la muestra. Cuando el nivel de significación se ha fijado y la varianza de la variable aleatoria es conocida (o se ha estimado) es posible controlar la potencia de la prueba manejando el tamaño muestral (o, en el caso de los diseños experimentales, manejando el número de repeticiones).

Curva de potencia

Hasta aquí se ha considerado solo la alternativa $\mu = 52$. Otras alternativas tendrán diferentes valores de β. Suponga de nuevo que se está contrastando la hipótesis $\mu = 50$ vs. $\mu \neq 50$ con $\sigma = 10$ y $\alpha = 0.05$, con una muestra de tamaño 25. Ahora, **si μ bajo la hipótesis alternativa es 53, se obtiene $\beta = 0.674$.**

El valor de β es más pequeño cuando las observaciones proceden de una población con $\mu = 53$ que cuando proceden de una población con $\mu = 52$. Para cualquier n y α fijos, se puede calcular β para una serie de valores de μ y graficar la llamada **curva de potencia** donde la potencia se define, como ya se indicara, $\pi = 1- \beta$.

La Figura 6.5 ilustra la función potencia $\pi(\mu)$ para una prueba bilateral.

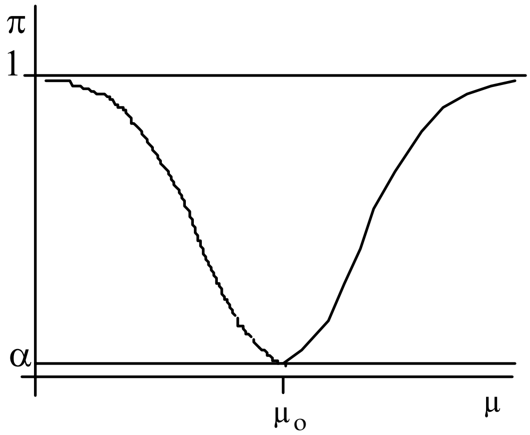


Figura 6.5: Curva de la función de potencia para una prueba bilateral.

Relación entre estimación por intervalo de confianza y prueba de hipótesis

En este capítulo y el anterior se han presentado dos procedimientos importantes en el marco de Inferencia Estadística: la *Estimación por Intervalos de Confianza* y la *Prueba de Hipótesis*.

Los Intervalos de Confianza se plantearon para estimar parámetros, mientras que las Pruebas de Hipótesis para tomar decisiones en relación a los valores postulados para ellos.

En muchos casos los Intervalos de Confianza y las Pruebas de Hipótesis se pueden utilizar alternativamente. Por ejemplo, en el caso de que se desee determinar si el rendimiento de una nueva variedad de soja es de 20 qq/ha, se plantean las siguientes hipótesis:

$$H_0: \mu = 20 \quad H_1: \mu \neq 20$$

Realizando la prueba correspondiente, se llegará a no rechazar o a rechazar H_0.

Este problema también se podría haber resuelto al obtener una estimación de μ por Intervalo de Confianza. Si el valor hipotético de μ (20 qq/ha) hubiera quedado comprendido dentro del intervalo no se habría rechazado la H_0 y en caso contrario se habría rechazado.

Finalmente, en el caso que se rechace H_0, se puede aplicar la estimación por intervalo para saber cúal es el valor del parámetro con una determinada confianza.

Ejercicios

Ejercicio 6.1

Una variable aleatoria sigue una distribución N(μ, 144) con μ desconocido.

a) ¿Se descartaría la hipótesis $\mu = 15$ en favor de la alternativa $\mu \neq 15$, para $\alpha = 0.05$, si una muestra aleatoria de n = 64 observaciones arroja una media igual a 20?

b) Construir un intervalo de confianza del 95% para μ.

c) Considerando la misma hipótesis del punto a), ¿qué sucedería con un nivel de significación del 1%?

d) Construir un intervalo de confianza del 99% para μ.

e) Probar H_0: $\mu = 15$ versus H_1: $\mu > 15$ para $\alpha = 0.05$ y $\alpha = 0.01$. Comparar con los resultados obtenidos en los puntos a) y c).

Ejercicio 6.2

Un proceso de fabricación produce 12.3 unidades por hora. Esta producción tiene una varianza igual a 4. Se sugiere un nuevo proceso que es costoso de instalar, pero se piensa que puede incrementar la producción. Para decidir si se hace el cambio o no, se prueban 10 máquinas nuevas y se observa que éstas producen en *promedio* 13.3 unidades.

a) Calcular la probabilidad del error de tipo II en la prueba para $\mu = 12.3$ vs $\mu > 12.3$ cuando la verdadera esperanza del nuevo proceso es $\mu = 14$. Trabajar con $\alpha = 0.01$.

Ejercicio 6.3

Al contrastar la hipótesis $\mu = 50$ vs $\mu > 50$, en una distribución normal con $\sigma = 2$,

b) ¿Con qué frecuencia sería aceptada esta hipótesis si la media verdadera fuese 51? Trabajar con $\alpha = 0.10$ y n = 10.

c) Si se desea mantener la probabilidad de Error Tipo I del punto anterior, ¿cómo se podría disminuir la frecuencia de aceptación de la hipótesis nula falsa?

Ejercicio 6.4

Un genetista afirma que el *rendimiento* de sus híbridos es distinto al de los progenitores, el cual es de 30 qq/ha. Si la desviación estándar es de 2 qq/ha y trabaja con una muestra de 10 híbridos:

¿Cuál es la probabilidad de que concluya que el rendimiento de los híbridos es igual al de los progenitores, si el rinde promedio es verdaderamente de 29 qq/ha? Trabajar con $\alpha = 0.05$.

Ejercicio 6.5

Se acepta que después de 3 años de almacenamiento el *vigor* de un arbusto forrajero medido como peso seco alcanzado a los 20 días de la germinación es de 45 mg promedio. Un nuevo método de almacenamiento se propone para aumentar el vigor. Se evalúan para ello 20 lotes de 10 semillas cada uno y al cabo de 3 años se las hace germinar, obteniéndose los siguientes resultados de peso seco promedio a los 20 días:

49	43	56	57	59	65	52	51	50	55
60	65	53	57	67	56	53	37	45	42

a) Plantear las hipótesis nula y alternativa asociadas al problema.

b) Realizar una prueba de hipótesis con un nivel de significación $\alpha = 0.01$.

c) De acuerdo a la conclusión que se obtuvo en el punto anterior, ¿se justifica realizar un cálculo de potencia?; ¿por qué?

Ayuda: si tuviera que calcular la potencia con la que se realizó la prueba, acepte la varianza muestral calculada como si se tratara de la varianza poblacional y tome a la media muestral como estimador de la verdadera media poblacional.

Ejercicio 6.6

Un tipo de ratón de laboratorio muestra una *ganancia media de peso* de 65 gr. durante los primeros tres meses de vida. Doce ratones fueron alimentados con una nueva dieta desde su nacimiento hasta los primeros tres meses de vida, observándose las siguientes ganancias de peso en gr.:

65	62	64	68	65	64	60	62	69	67	62	71

a) ¿Hay razón para creer que la dieta produce una variación significativa en la cantidad de peso ganado? Trabajar con $\alpha = 0.05$.

b) Calcular para la prueba planteada, las potencias para diferentes valores de μ_1 variando en el intervalo [62 gr., 70 gr.] y dibujar la curva de potencia.

7

Inferencia Sobre la Esperanza y la Varianza de Variables Aleatorias Distribuidas Normalmente

Introducción

Como se recordará de los Capítulos 5 y 6, los intervalos de confianza se plantearon para dar una medida de confianza a la estimación de parámetros, mientras que las pruebas de hipótesis para tomar decisiones con relación a los valores postulados para los mismos. En este capítulo se abordará esta temática, relacionada con las pruebas de hipótesis y la construcción de intervalos de confianza para la esperanza y varianza de una y dos distribuciones normales y se analizarán situaciones en donde se pueden aplicar estas metodologías.

Prueba de hipótesis acerca de una esperanza

Caso 1: Se conoce la varianza σ^2

Ejemplo 7.1

La producción media de trigo por hectárea en una región es de 2200 kg. con una desviación estándar (σ) de 450 kg. Se desea establecer si la aplicación de fertilizantes modifica el rendimiento medio del trigo.

De acuerdo con el algoritmo presentado en el capítulo anterior se tiene:

a) Se establecen las hipótesis:

$$H_0: \mu \leq 2200 \text{ kg} \;\; \text{y} \;\; H_1: \mu > 2200 \text{ kg}.$$

b) El ensayo consiste en elegir 20 has. (una en cada chacra de la región) en forma aleatoria y fertilizarlas, evaluando su rendimiento a cosecha.

c) Usando el estadístico Z = $\left(\dfrac{\overline{X}-\mu_0}{\dfrac{\sigma}{\sqrt{n}}}\right)$ donde Z ~ N (0,1) bajo H_0 y

d) Eligiendo $\alpha = 0.05$

e) Se determinan los límites de la región de aceptación: en este caso se trata de una prueba unilateral derecha cuyo único punto crítico es $Z_{1-\alpha} = 1.645$, el cual se obtiene de la Tabla de Distribución Normal Estándar.

f) Según el experimento planificado en el paso b), la producción media obtenida fue de 2650 kg.

g) Calculando el estadístico:

h) $Z = \dfrac{2650 - 2200}{450/\sqrt{20}} = 4.47$

Como puede verse, Z = 4.47 > 1.645 por lo que se rechaza H_0 y se concluye que la producción media de trigo por ha. con fertilización, en la región, es significativamente mayor que 2200 kg.

La Figura 7.1 muestra las zonas de aceptación y rechazo de H_0 en una prueba unilateral derecha.

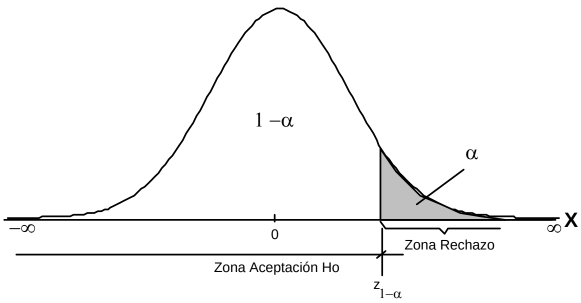


Figura 7.1: Región crítica para un contraste unilateral derecho, cuyo estadístico tiene distribución normal estándar bajo H_0 .

Para este problema el intervalo de confianza apropiado es un intervalo unilateral izquierdo, ya que lo que importa es dar un límite inferior para el rendimiento cuando se usa fertilización. Este límite esta basado en:

$$P\left(\overline{X} - z_{1-\alpha}\sqrt{\sigma^2/n} < \mu\right) = (1-\alpha)$$

En este caso el límite inferior resultante es $2650 - 1.645\sqrt{450^2/20} = 2484$, por lo que se espera que el rendimiento del trigo fertilizado no será menor que 2484, con una confianza del 95%.

Ejemplo 7.2

Considérese ahora que se desea probar si una nueva técnica siembra en vivero produce un aumento de la longitud, en plantines de algarrobo, al cabo de tres meses de realizada la siembra. Bajo la técnica tradicional, los plantines alcanzan una altura promedio de 15 cm y por experiencias previas se espera que la nueva técnica produzca un incremento de la longitud. Para evaluar el nuevo procedimiento se proponen las siguientes hipótesis estadísticas: H_0: $\mu = 15$ cm. vs H_1: $\mu > 15$ cm.

El ensayo consiste en evaluar la altura de 16 plantines de algarrobo al cabo de 3 meses desde la siembra.

El estadístico propuesto fue Z~N(0,1) ya que se conoce la varianza de la distribución ($9cm^2$) y la variable altura de plantines se supone normal. El nivel de significación elegido fue $\alpha = 0.05$ y la región de aceptación resultante fue $(-\infty, 1.645)$. Esta es una prueba unilateral derecha y el valor obtenido del estadístico fue 1.333, con lo cual **no se rechazó H_0.**

La pregunta que puede derivarse de este resultado es ¿qué chance de rechazar H_0 se tenía en la prueba anterior, si la verdadera esperanza de la distribución de alturas era efectivamente 17 cm?

Para responder a esta pregunta lo que se debe hacer es calcular el valor de la función potencia de la prueba para ese valor de 17 cm. Esta evaluación requiere el cálculo de la probabilidad de Error de Tipo II ya que la potencia se define como 1-P(Error Tipo II). Como se indicó en el capítulo anterior, probabilidad de Error de Tipo II, para una prueba unilateral derecha, está dada por la siguiente expresión:

$$\beta = P\left(Z \leq Z_{1-\alpha} - \frac{\mu - \mu_0}{\sqrt{\frac{\sigma^2}{n}}} \right)$$

donde Z representa al estadístico estandarizado de la prueba, que se distribuye N (0,1). Luego, para el problema planteado:

$$\beta = P (Z \leq 1.645 - 8/3) = P (Z \leq - 1.022) = 0.15339$$

Este resultado indica que la probabilidad de aceptar H_0 siendo falsa es 0.15, lo cual da una potencia (1 - β) de 0.85. En consecuencia, el no rechazo de H_0 en vista de la alta potencia la prueba, es confiable.

Caso 2: No se conoce la varianza σ^2

Ejemplo 7.3

Se piensa que la producción promedio de un nuevo cultivar de trigo es superior al rendimiento promedio del trigo que se siembra usualmente, que es de 2000 kg./ha. Para establecer si esto es cierto se procede a realizar una prueba de hipótesis.

De acuerdo al algoritmo presentado en el capítulo anterior se tiene:

a) Sea μ la esperanza de la distribución de rendimientos en la región con el nuevo cultivar, el problema consiste en decidir entre:

$$H_0: \mu \leq 2000 \text{ kg./ha} \;\; \text{y} \;\; H_1: \mu > 2000 \text{ kg./ha}$$

b) Para probar esta hipótesis se seleccionan aleatoriamente, dentro de la región de interés, 12 campos de 5 has. cada uno en los que se sembrará el nuevo cultivar, registrándose su rendimiento a cosecha.

c) Se supone que el rendimiento promedio es una variable aleatoria normal ya que a través del teorema central del límite puede justificarse que la variable rendimiento, por ser en este ejemplo acumulación de miles de rendimientos individuales, tiene esta distribución.

Se propone el estadístico $T = \left(\dfrac{\bar{X} - \mu_0}{\frac{S}{\sqrt{n}}} \right)$ que bajo H_0 se distribuye como una T de Student con n-1 grados de libertad. Recuérdese que esto se debe a que se desconoce la varianza de la distribución y se la estima por S^2.

d) Se fija un nivel de significación $\alpha = 0.05$

e) Como se espera que el nuevo cultivar tenga un rendimiento promedio más alto, cuanto mayor resulte la media muestral, más se alejará (hacia la derecha) el estadístico T de su valor esperado bajo la hipótesis nula que es 0. Luego, se está en presencia de un contraste unilateral derecho cuya región de rechazo queda definida por el intervalo $(T_{(n-1);\ 1-\alpha}, \infty)$. La probabilidad de esta región bajo H_0 es, obviamente, α. Luego el punto crítico es $T_{(12-1)\ ;\ 0.95} = 1.796$, el cual se obtiene de la Tabla T de Student.

f) El rendimiento promedio del nuevo cultivar calculado a partir de las 12 parcelas es $\bar{X}$ = 2020 y la desviación estándar estimada S = 100.

g) Luego el valor del estadístico es:

$$T = \frac{2020 - 2000}{100/\sqrt{12}} = 0.692$$

Dado que T = 0.692 < $T_{(12-1)\,;\,0.95}$ = 1.796 no se rechaza H_0. Se concluye que no hay evidencia de que el nuevo cultivar tenga un rendimiento promedio mayor a 2000 kg./ha.

La construcción de un intervalo de confianza bilateral para μ sería útil para dar un informe más completo. Dicho intervalo está dado por la siguiente expresión:

$$\boxed{P\left(\bar{X} - T_{1-\alpha/2}\sqrt{S^2/n} < \mu < \bar{X} + T_{1-\alpha/2}\sqrt{S^2/n}\right) = (1-\alpha)}$$

Prueba de hipótesis acerca una varianza

Ejemplo 7.4

Una firma agroindustrial desea incorporar un nuevo mecanismo en las máquinas enfardadoras que fabrica. El ingeniero a cargo del proyecto sospecha que esta innovación puede producir un aumento de la varianza del peso de los fardos. La desviación estándar que se obtiene con la maquinaria sin modificar es de 1.5 kg. Para evaluar el nuevo mecanismo, se realizó un ensayo tomando 10 fardos al azar de un lote de alfalfa. Los pesos de dichos fardos fueron: 28.3; 27.8; 29.3; 30.1; 32.5; 27.2; 25.3; 32.2; 33.6; 30.7, con varianza muestral = 6.87.

3. Con esta evidencia se desea probar la siguiente hipótesis:

   $H_0{:}\sigma^2 = 2.25$ vs. $H_1{:}\sigma^2 > 2.25$

4. El estadístico a utilizar es $\frac{(n-1)S^2}{\sigma^2}$ que se distribuye como $\chi^2_{(n-1)}$.

5. Se fija α = 0.10.

6. Luego, se obtiene el cuantil 0.90 de la distribución $\chi^2_{(10-1)}$ en la Tabla

Chi-cuadrado, que es $\chi^2_{(9;\ 0.90)}$ = 14.68, delimitando así las regiones de no rechazo y rechazo de H_0.

7. Se evalúa el estadístico $(n - 1)\ S^2 / \sigma_0^2 = 27.48$
8. Dado que 27.48 está dentro de la región de rechazo, se rechaza H_0. Esto implica que el nuevo mecanismo provoca un aumento de la varianza del peso de los fardos.

Nota: en el caso que se desee contrastar una alternativa bilateral:

$$H_0: \sigma^2 = \sigma_0^2 \text{ y } H_1: \sigma^2 \neq \sigma_0^2$$

se usará igualmente el estadístico $\chi^2 = (n - 1)\ S^2 / \sigma_0^2$ para probar la hipótesis nula y se determinará la región crítica, que estará delimitada por los cuantiles $\alpha/2$ y $1-\alpha/2$ de la distribución chi-cuadrado.

Estimación por Intervalo de una varianza

Si S^2 es el estimador de σ^2 para muestras de tamaño n, entonces se puede proponer como función para construir el intervalo a la siguiente expresión:

$$g(\sigma^2, S^2) = (n-1)\ S^2 / \sigma^2$$

Se sabe que $(n-1)\ S^2/\sigma^2 \sim \chi^2_{n-1}$

Siendo los cuantiles $q_1 = \chi^2_{(n-1)\,;\,(\alpha/2)}$ y $q_2 = \chi^2_{(n-1)\,;(1-\alpha/2)}$, se tiene:

$$P(q_1 \leq S^2 (n-1) / \sigma^2 \leq q_2) = 1 - \alpha$$

Luego, despejando σ^2:

$$P(S^2 (n-1) / q_1 \geq \sigma^2 \geq S^2 (n-1) / q_2) = 1 - \alpha$$

Reordenando se tiene:

$$\boxed{P(S^2 (n-1) / q_2 \leq \sigma^2 \leq S^2 (n-1) / q_1) = 1-\alpha}$$

Luego: **LI** $= S^2 (n-1) / q_2$ y **LS** $= S^2 (n-1) / q_1$ son los límites inferior y superior, respectivamente, del intervalo de confianza $1-\alpha$ para σ^2.

Ejemplo 7.5

Retomando el ejemplo anterior, si se quiere estimar por intervalo de confianza al 90% la varianza del peso de los fardos se tendrá:

LI= 6.87 (9)/ 16.92 = 3.25 y LS= 6.87 (9)/ 3.32 =22.9

Prueba de hipótesis para dos varianzas

A veces se quiere comparar las varianzas de dos variables aleatorias con distribución normal. Para ello usualmente se cuenta con muestras independientes, una de cada distribución, de tamaños n_1 y n_2 respectivamente.

Si las varianzas de las poblaciones que se están muestreando son iguales, entonces el cociente S_1^2/S_2^2 se distribuye como una distribución F con (n_1-1) y (n_2-1) grados de libertad.

Definición 7.1: Distribución F

Sean X_1 y X_2 variables aleatorias (independientes) con distribución χ^2 con ν_1 y ν_2 grados de libertad respectivamente y sea F la variable aleatoria construida a partir de las primeras mediante la siguiente expresión:

$$F = \frac{\frac{X_1}{\nu_1}}{\frac{X_2}{\upsilon_2}}$$

luego F se distribuye como una F de Snedecor con ν_1 y ν_2 grados de libertad.

Denotaremos a una distribución F con ν_1 y ν_2 grados de libertad como $F_{\nu_1;\nu_2}$.

Esta distribución, atribuida a Snedecor, está definida para valores no negativos y se caracteriza por los grados de libertad del numerador y del denominador de la expresión anterior. La distribución es asimétrica y el grado de asimetría depende los grados de libertad.

Un resultado importante es que si se toman dos muestras aleatorias e independientes de tamaños n_1 y n_2 de una distribución normal con varianza σ^2 y a partir de ellas se calculan las varianzas muestrales S_1^2 y S_2^2 se tiene que S_1^2 $(n_1-1) / \sigma^2 \sim \chi^2$ con (n_1-1) grados de libertad y S_2^2 $(n_2-1) / \sigma^2 \sim \chi^2$ con (n_2-1) grados de libertad y usando la Definición 7.1, se puede justificar que:

$$\frac{S_1^2}{S_2^2} = \frac{\dfrac{S_1^2\ (n_1 - 1)/\sigma^2}{(n_1 - 1)}}{\dfrac{S_2^2\ (n_2 - 1)/\sigma^2}{(n_2 - 1)}} \sim F_{(n_1-1)(n_2-1)}$$

La “Tabla de Cuantiles de la Distribución F” del Anexo presenta algunos cuantiles correspondientes a la distribución F de Snedecor acumulada, para varias combinaciones de grados de libertad del numerador y del denominador del cociente dado en la definición. Por ejemplo, si $F \sim F_{3,10}$ entonces $P[F \leq 4.83] = 0.975$.

Ejemplo 7.6

Retomando el ejemplo 4.3, del Capítulo 4; cuyos estadísticos muestrales se reproducen en la siguiente tabla:

Lote	n	Media	Varianza
1	10	46.3	39.4
2	10	52.6	63.8

las varianzas muestrales de ambos lotes parecen diferentes. ¿Pero es ésta diferencia significativa? Desde el punto de vista estadístico se puede presentar esta cuestión en términos de probabilidad.

Si las varianzas de estas poblaciones fueran iguales entonces el cociente de las varianzas muestrales se distribuye según una F con 9 y 9 grados de libertad. Por lo tanto el 95% de todos los valores muestrales de los cocientes de pares de varianzas obtenidos con tamaños muestrales de 10 estarán comprendidos entre los percentiles 0.025 y 0.975 de esa distribución. En este caso estos cuantiles asumen los valores 0.2484 y 4.0260 respectivamente. Luego como el cociente 63.8 / 39.4 = 1.62 está entre 0.2484 y 4.0260, se puede decir que el cociente observado es compatible con la suposición de que las **varianzas son iguales** y en consecuencia no hay evidencia para rechazar la hipótesis de igualdad de varianzas.

Prueba de hipótesis y estimación por intervalo de confianza para la diferencia de dos esperanzas

Caso 1: Las varianzas son conocidas

Sean μ_1 y μ_2 las esperanzas de las distribuciones 1 y 2 respectivamente y $\bar{X}_1$ y $\bar{X}_2$ dos estimadores independientes de las respectivas esperanzas.

Como las medias muestrales $\bar{X}_1$ y $\bar{X}_2$ tienen distribución normal con parámetros (μ_1, σ_1^2/n_1) y (μ_2, σ_2^2/n_2) respectivamente, y como la diferencia de variables aleatorias normales e independientes es también una variable aleatoria normal con esperanza igual a la diferencia de las esperanzas y varianza igual a la suma de las varianzas, se tiene:

$$(\bar{X}_1 - \bar{X}_2) \sim N((\mu_1-\mu_2), (\sigma_1^2/ n_1 + \sigma_2^2/ n_2))$$

de lo que se deduce que : $Z = \dfrac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}} \sim N(0,1)$

donde σ_1^2 y σ_2^2 corresponden a las varianzas de las distribuciones y n_1 y n_2 a los tamaños de las muestras a partir de las cuales se calcularon $\bar{X}_1$ y $\bar{X}_2$.

Esta es la expresión del estadístico Z que se usa para probar hipótesis referidas a la diferencia entre las medias, siguiendo el procedimiento general presentado en el Capítulo 6.

El intervalo de confianza para ($\mu_1 - \mu_2$) con una confianza ($1 - \alpha$), se construye en forma similar al de la esperanza de una distribución con $q_1= Z_{(\alpha/2)}$ y $q_2= Z_{(1-\alpha/2)}$. Luego:

$$P\left(Z_{(\alpha/2)} \leq \left(\frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}}\right) \leq Z_{(1-\alpha/2)}\right) = 1-\alpha$$

despejando ($\mu_1 - \mu_2$) y teniendo en cuenta que $Z_{(1-\alpha/2)} = - Z_{(\alpha/2)}$ se tiene:

$$P\left((\bar{X}_1-\bar{X}_2) - Z_{(1-\alpha/2)} \cdot \sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}} \le \mu_1 - \mu_2 \le (\bar{X}_1-\bar{X}_2) + Z_{(1-\alpha/2)} \cdot \sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}} \right) = 1-\alpha$$

Así: **LI** $= (\bar{X}_1-\bar{X}_2) - Z_{(1-\alpha/2)} \cdot \sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}$ y **LS** $= (\bar{X}_1-\bar{X}_2) + Z_{(1-\alpha/2)} \cdot \sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}$ son los límites inferior y superior, respectivamente, del intervalo de confianza 1-α para μ_1-μ_2.

Caso 2: Las varianzas son desconocidas

Para hacer inferencia sobre las esperanzas de dos distribuciones normales cuando no se conocen las varianzas, es necesario establecer previamente si dichas varianzas son o no iguales ya que de ello depende el estadístico a usar en la prueba de hipótesis como, así también, la función con la que se construye el intervalo de confianza. Por lo tanto, antes de probar la diferencia entre dos medias, se prueba si las varianzas poblacionales son iguales.

Caso 2-a: Las varianzas son desconocidas e iguales

Ejemplo 7.7

Suponga que se quieren comparar dos variedades de maní, en cuanto al contenido de aceite de las semillas. Las hipótesis de esta prueba son H_0: $\mu_1 = \mu_2$ vs H_1: $\mu_1 \neq \mu_2$.

Para probar las hipótesis anteriores se diseña un ensayo en el que para cada variedad se obtienen los contenidos de aceite de 10 bolsas de 1kg de semillas de maní cada una extraídas aleatoriamente de un semillero. Los resultados del ensayo son los siguientes:

Tabla 7.1: Resultados de un ensayo comparativo de contenido de aceite en la semilla de maní de dos variedades

Variedad	n	$\bar{X}$	S^2
1	10	160.4	65.3
2	10	165.6	67.9

La prueba de esta hipótesis se realiza con un estadístico que depende de la igualdad de varianzas. Luego, para elegir el estadístico de la prueba se debe probar la hipótesis

H_0: $\sigma_1^2 = \sigma_2^2$ vs H_1: $\sigma_1^2 \neq \sigma_2^2$. Para ello se utiliza el estadístico $F=\dfrac{S_1^2}{S_2^2}$ que bajo H_0 se distribuye como F con 9 y 9 grados de libertad. Por lo tanto la región de aceptación para un nivel de significación del 5% está delimitada por los valores 0.248 y 4.03, correspondientes a los cuantiles $\alpha/2$ y $(1 - \alpha/2)$, respectivamente.

Calculando el estadístico propuesto se obtiene F=0.962 que está dentro de la región de aceptación. Luego, no se rechaza la igualdad de varianzas y en consecuencia el estadístico apropiado para la prueba de hipótesis de igualdad de medias es el siguiente:

$$T = \frac{\left(\bar{X}_1 - \bar{X}_2\right)}{\sqrt{\dfrac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1+n_2-2}\left(\dfrac{1}{n_1}+\dfrac{1}{n_2}\right)}}$$

que se distribuye según una *T* de Student con $(n_1 + n_2 - 2)$ grados de libertad. Fijando α=0.01 se tiene que los puntos críticos que delimitan la región de aceptación son -2.878 y 2.878.

Utilizando los valores muestrales de las medias y varianzas se obtiene T = -1.43. Como este valor está dentro de la región de aceptación, se concluye que no hay evidencia para rechazar H_0.

Los argumentos propuestos para la construcción del intervalo de confianza son similares a los ya señalados.

Luego, para encontrar por ejemplo el intervalo de confianza al 90% con tamaños muestrales n_1 = 10 y n_2 = 10 se necesitan los cuantiles $q_1 = T_{18;\ (0.05)} = -1.734$ y $q_2 = t_{18;\ (0.95)} = 1.734$ y, siguiendo la metodología general propuesta, se tiene:

$$\boxed{P\left((\bar{X}_1 - \bar{X}_2) - 1.730\sqrt{\left(S_p^2 \cdot \frac{n_1+n_2}{n_1 \cdot n_2}\right)} \le \mu_1 - \mu_2 \le (\bar{X}_1 - \bar{X}_2) + 1.734\sqrt{\left(S_p^2 \cdot \frac{n_1+n_2}{n_1 \cdot n_2}\right)}\right)=0.90}$$

por lo cual:

$LI = (\bar{X}_1-\bar{X}_2) - 1.734 \cdot \sqrt{\left(S_p^2 \cdot \dfrac{n_1+n_2}{n_1 \cdot n_2}\right)}$ y $LS = (\bar{X}_1-\bar{X}_2) + 1.734 \cdot \sqrt{\left(S_p^2 \cdot \dfrac{n_1+n_2}{n_1 \cdot n_2}\right)}$ son los limites de confianza para este ejemplo.

Caso 2-b: Las varianzas son desconocidas y diferentes

Recuérdese que según lo estudiado en el Capítulo 4, la desviación estándar de la diferencia de medias muestrales, $S_{\bar{X}_1-\bar{X}_2}$, se calcula como:

$$S_{\bar{X}_1-\bar{X}_2} = \sqrt{\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}}$$

La "estandarización" que se obtiene utilizando las estimaciones de las varianzas muestrales es la siguiente:

$$T' = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{S_{\bar{x}_1 - \bar{x}_2}}$$

que tiene distribución T de Student con los grados de libertad que se especifican a continuación:

$$\nu = \frac{\left(S_1^2/n_1 + S_2^2/n_2 \right)^2}{\frac{\left(S_1^2/n_1\right)^2}{n_1+1} + \frac{\left(S_2^2/n_2\right)^2}{n_2+1}} - 2$$

Luego, la prueba de hipótesis utiliza el estadístico T' y el proceso de construcción del intervalo de confianza tiene la expresión final que se presenta a continuación.

$$\boxed{P\left((\bar{X}_1 - \bar{X}_2) - T_{(\nu\,;\,1-\alpha/2)} \cdot S_{\bar{X}_1-\bar{X}_2} \leq \mu_1 - \mu_2 \leq (\bar{X}_1 - \bar{X}_2) + T_{(\nu\,;\,1-\alpha/2)} \cdot S_{\bar{X}_1-\bar{X}_2} \right) = 1 - \alpha}$$

Caso 3: Dos muestras no independientes

Se analizará ahora la diferencia entre las esperanzas de dos grupos cuando los datos se obtienen de muestras que están **relacionadas**; es decir, los resultados del primer grupo no son independientes de los del segundo. Por ejemplo, esto ocurre cuando se mide la presión arterial en cada uno de los individuos de un grupo experimental antes y después de la administración de una droga. El objetivo es comprobar si la droga produce efectos en la presión sanguínea. Los pares de observaciones (antes y después) obtenidas en cada individuo **no son independientes** ya que la presión arterial posterior a la administración de la droga depende de la presión arterial inicial.

Una situación equivalente ocurre cuando, por ejemplo, se desea probar si hay diferencias en el tamaño y calidad de las semillas recolectadas de flores de la parte apical y basal en plantas de alfalfa. En este caso se puede señalar que la vinculación o dependencia entre las observaciones, sobre flores de la parte apical y basal, está relacionada con la calidad de la planta madre. Así, se encontrarán plantas que producen semillas de alta calidad y otras con semillas de baja calidad, independientemente de las variaciones entre las partes alta y baja de la planta.

Otro ejemplo: supóngase que en un ensayo para comparar rendimientos con dos fertilizantes "A" y "B" se siembran diez parcelas de trigo tratadas con el fertilizante "A" en 10 zonas experimentales y otras 10 parcelas de trigo tratadas con el fertilizante "B" en otras 10 zonas experimentales. Si el promedio de las parcelas tratadas con el fertilizante "A" se compara con el obtenido para el fertilizante "B", parte o gran parte de la diferencia observada (si la hay) puede deberse a los diversos tipos de terreno o a las distintas condiciones climáticas de las zonas experimentales, estas *fuentes de variación* pueden enmascarar o confundir el efecto diferencial de los fertilizantes que se comparan. Un arreglo diferente de este experimento, que permite disminuir las fuentes de variación indeseables, consiste en obtener **observaciones apareadas**. Este **diseño de experimento** alternativo podría ser el siguiente: se eligen al azar 10 zonas experimentales y en cada una de ellas se siembran dos parcelas contiguas (para asegurar que las condiciones locales sean las mismas) y cada una es tratada con el fertilizante "A" o "B" respectivamente. En este caso, también se obtienen 20 observaciones pero éstas están *apareadas* de a dos.

Con el término **observaciones apareadas** se hace referencia al diseño de experimentos que produce observaciones "de a pares" de las dos distribuciones que se comparan. En este tipo de diseño la variable de interés es la **diferencia** entre los valores de cada uno de los pares observados. El objetivo es reducir la variabilidad debida a factores que introducen efectos extraños a aquel que se desea medir.

Sea X_{i1} el primer miembro del par i-ésimo y X_{i2} el segundo miembro, para n pares de observaciones se tendrá: (X_{11},X_{12}) , (X_{21},X_{22}) , (X_{31},X_{32}) , ... , (X_{n1},X_{n2}). Si se toman las diferencias $d_i = X_{i1}- X_{i2}$, se tendrá un conjunto de n observaciones, cada una de las cuales es una diferencia entre dos observaciones originales.

El uso de este diseño es recomendable cuando se desea eliminar una fuente de variación que tiene un efecto aditivo sobre ambos miembros del par. Es decir, se supone que el efecto es esencialmente el de aumentar o disminuir, mediante alguna constante, cada una de las esperanzas de modo que al tomar la diferencia entre los miembros del par se elimine dicho efecto.

El uso arbitrario de este diseño, cuando el apareamiento no implica una disminución de las fuentes de variación no deseadas, produce una perdida de potencia.

Prueba T para observaciones apareadas

Esta prueba se basa en la distribución de la variable diferencia entre los pares de observaciones. Si X_{i1} y X_{i2} tienen distribución normal, entonces, las $d_i = X_{i1} - X_{i2}$ tendrán distribución normal con esperanza $\mu_d = \mu_1 - \mu_2$ y varianza σ_d^2. El estimador de $\mu_1 - \mu_2$ es $\bar{d} = \frac{1}{n}\sum_{i=1}^{n} d_i$ y el estimador de σ_d es $S_d = \sqrt{\frac{\sum_{i=1}^{n}\left(d_i - \bar{d}\right)^2}{n-1}}$

Nota: Por la falta de independencia entre los elementos de los pares de observaciones, la varianza de la diferencia es menor que la suma de las varianzas de las variables originales, de allí la ventaja de este diseño.

Si la hipótesis nula que se quiere probar es $\mu_1 - \mu_2 = 0$, esto implica $\mu_d = 0$, luego para probar esta hipótesis el estadístico apropiado es:

$$T = \frac{\bar{d}}{S_d / \sqrt{n}} \sim T_{(n-1)}$$

donde el n es el número de pares de observaciones en la muestra.

Nota: Esta prueba no requiere el supuesto usual de homogeneidad de las varianzas σ_1^2 y σ_2^2 ya que se basa en la varianza de las diferencias que se estima independientemente de éstas.

Para la construcción del intervalo de confianza correspondiente, se sigue el procedimiento conocido y el intervalo se basa en la siguiente expresión:

$$\boxed{P\left(\bar{d} - T_{(n-1);(1-\alpha/2)}\frac{S_d}{\sqrt{n}} < \mu_1 - \mu_2 < \bar{d} + T_{(n-1);(1-\alpha/2)}\frac{S_d}{\sqrt{n}}\right) = (1-\alpha)}$$

Ejemplo 7.8

Se quiere comparar el efecto de dos virus sobre plantas de tabaco. Para esto se realizó el siguiente experimento:

Se seleccionaron al azar 8 plantas y en cada una de ellas se tomaron 2 hojas apicales. Sobre cada una de ellas se aplicaron los preparados conteniendo los virus cuyos efectos se querían evaluar. La variable de respuesta fue la superficie en mm^2 de las lesiones locales que aparecían como pequeñas manchas oscuras en las hojas. Los resultados fueron:

Preparado 1	Preparado 2	d_i
31	18	13
20	17	3
18	14	4
17	11	6
9	10	-1
8	7	1
10	5	5
7	6	1
$\bar{X}_1=15$	$\bar{X}_2=11$	$\bar{d}=4$

Como el objetivo del ensayo fue comparar si existían diferencias entre los efectos de los dos virus se planteó la siguiente hipótesis: H_0: $\mu_1 = \mu_2$ vs H_1: $\mu_1 \neq \mu_2$ usando como estadístico:

$$T = \frac{\bar{d}}{S_d / \sqrt{n}}$$

que se distribuye como una T de Student con (8-1=7) grados de libertad.

Fijando $\alpha = 0.05$ se determina la región de aceptación como el intervalo ($t_{\alpha/2}$= -2.365 , $t_{1-\alpha/2}$ = 2.365). Luego, evaluando el estadístico se obtiene:

$t = \frac{4}{4.30/\sqrt{8}} = 2.63 > 2.365$ por lo que se rechaza H_0.

De acuerdo al resultado anterior, se concluye que las diferencias observadas entre las áreas dañadas por uno u otro virus son estadísticamente significativas.

Nota: Como siempre, se pueden derivar alternativas unilaterales para esta prueba.

Ejercicios

Ejercicio 7.1

Se considera que la fibra de un tipo de algodón es de buena calidad si su longitud media es mayor a 210 mm, con una desviación estándar de 50 mm. Para saber si un lote cumple con las especificaciones se toman 50 bolsas y de cada una de ellas se extraen 100 fibras y se calcula *la longitud promedio por bolsa*.

a) ¿Se trata de una prueba bilateral, unilateral derecha, o unilateral izquierda?

b) ¿Cuál es el promedio de 50 bolsas más pequeño para que un lote sea aceptado si se trabaja con un nivel de significación del 5%?

Ejercicio 7.2

Cuando la cantidad de semillas de soja que quedan en el suelo luego de pasar la cosechadora es igual o mayor a 80 semillas/m^2, la pérdida de producción, en qq/ha, es grande. Un productor decide probar el funcionamiento de su máquina y para ello luego de cosechar una parcela cuenta en 10 unidades de 1 m^2 cuántas semillas quedan en el suelo. Los resultados fueron, en semillas/m^2:

77 73 82 82 79 81 78 76 76 75

a) ¿Se puede concluir, trabajando con un nivel de significación del 10%, que la cosechadora está funcionando bien?, es decir, ¿está la perdida dentro de los límites admisibles?

b) Construir un intervalo de confianza para μ apropiado para el problema.

Ejercicio 7.3

Referido al problema anterior:

a) Si las normas técnicas indican que la desviación estándar del número de semillas caídas por m^2 no debería ser superior a 5, ¿qué se debería concluir sobre la máquina trabajando con un nivel de significación $\alpha = 0.10$?

b) Construir un intervalo de confianza para σ^2.

Ejercicio 7.4

Un experimentador avícola considera que al suministrar una ración especial a pollitos de la raza Cornich, ha de lograr un peso medio superior a 700 gr. por animal luego de cuatro semanas de alimentación. Para verificarlo alimenta con la ración a un lote de 50 pollitos y a los 28 días obtiene un peso promedio de 730 gr. con una desviación estándar de 40.21 gr.

a) Establecer las hipótesis nula y alternativa.

b) Realizar la prueba correspondiente utilizando $\alpha = 0.05$.

c) Construir un intervalo de confianza para μ.

Ejercicio 7.5

Para evaluar la homogeneidad de la fertilidad de un suelo se tomaron alícuotas de 20 extracciones de suelo y se midió su contenido de nitrógeno. Los resultados, en ppm, fueron:

0.50	0.48	0.39	0.41	0.43	0.49	0.54	0.48	0.52	0.51
0.49	0.47	0.44	0.45	0.40	0.38	0.50	0.51	0.52	0.45

Se acepta que un suelo es homogéneo en fertilidad, si el contenido de nitrógeno presenta una varianza de a lo sumo 0.005.

Con los datos de la muestra, construir un intervalo de confianza apropiado (unilateral o bilateral) al 90 % y evaluar a partir de él si el suelo es homogéneo o no en su fertilidad.

Ejercicio 7.6

Los siguientes datos corresponden a los residuos de Parathion (en ppm.) en plantas de un lote de apio. Los resultados obtenidos fueron:

0.26	0.52	0.52	0.50	0.45	1.08	0.34	0.33	0.25	0.29	0.18	0.42	0.15	1.05
0.95	0.92	0.52	0.41	0.77	0.44	0.29	0.44	0.64	0.36	0.50	0.60	0.92	0.58
0.46	0.52	0.24	0.53	0.39	0.40	0.54	0.47	0.43	0.32	0.38	0.31	0.25	0.60
0.84	0.55	0.26	0.51	0.50	0.75	0.54	0.60	0.71	0.56	0.52	0.49	0.50	0.43
0.59	0.26	0.24	0.66	0.66	0.56	0.66	0.92	0.67	0.52	0.36	0.50	0.52	0.45
0.92	0.51	0.40	0.60	0.85	0.53	0.44	0.30						

Un ente fiscalizador establece que si el residuo de insecticida es mayor que 0.50 ppm, se debe rechazar el lote de plantas de apio para consumo humano. ¿Qué decisión se tomaría, a partir de esta información, trabajando con $\alpha = 0.01$?

Ejercicio 7.7

Uso de la tabla de la Distribución F de Snedecor.

La tabla que se presenta en el Anexo muestra algunos cuantiles correspondientes a la distribución F acumulada para varias combinaciones de grados de libertad del numerador y del denominador. Como ejemplo del uso de la tabla, supóngase que se quiere encontrar la probabilidad de que una variable cuya distribución es F con 3 y 10 grados de libertad tome valores menores o iguales a 4.83. Esto es P ($F_{3,10} \leq 4.83$). Para hallar esta probabilidad se busca en la hoja de la tabla (notar que la misma ha sido fraccionada en varias hojas) en cuyo vértice superior izquierdo aparece un 3 (*grados de libertad del numerador*). Luego, sobre el margen izquierdo se localiza la fila que comienza con el número 10 y que corresponde a los *grados de libertad del denominador* de la distribución F. En la fila seleccionada, se busca 4.83. El valor que encabeza la columna donde se encuentra 4.83 es 0.975, luego P ($F_{3,10} \leq 4.83$) = 0.975; es decir 4.83 es el cuantil 0.975 de una distribución F de Snedecor con 3 y 10 grados de libertad.

Como ejercicio sobre el uso de esta tabla, encuéntrese:

a) P ($F \leq 1.8376$) si F se distribuye con distribución $F_{20,11}$.

b) El cuantil 0.10 de una distribución $F_{15,12}$.

c) El valor de una variable distribuida como una $F_{1,5}$ que acumula el 95% de los valores de la distribución.

Ejercicio 7.8

Un grupo de conejos fue sometido a una serie de situaciones de tensión que producían una respuesta de temor. Después de un período de tiempo bajo estas condiciones, los conejos fueron comparados con los de un grupo control, que no había sido sometido a tensión. La variable de respuesta fue el peso (en mg) de la glándula suprarrenal. Los resultados fueron:

Grupo Experimental:	3.8	6.8	8.0	3.6	3.9	5.9	6.0	5.7	5.6	4.5	3.9	4.5
Grupo Control:	4.2	4.8	4.8	2.3	6.5	4.9	3.6	2.4	3.2	4.9		

a) Comparar el peso de la glándula suprarrenal entre el grupo control y el experimental con un nivel de significación del 5%.

b) Construir un intervalo de confianza para la diferencia de medias poblacionales.

c) ¿Qué supuestos se necesitan para que los procedimientos utilizados en a) y b) sean estadísticamente válidos?

Ejercicio 7.9

Se está experimentando con un herbicida en maíz, y para ponerlo a prueba se evalúan los rendimientos de 12 parcelas experimentales. En 6 de ellas se utilizó el nuevo herbicida y en las restantes un herbicida tradicional como control. Los resultados del ensayo, expresados en quintales por hectárea, son los siguientes:

Nuevo herbicida:	68.1	74.6	64.4	69.2	61.8	57.9
Viejo herbicida:	64.7	62.5	66.8	69.2	53.9	58.5

a) ¿Qué se puede decir del desempeño del nuevo herbicida en relación al control, trabajando con un nivel de significación $\alpha = 0.10$?

b) ¿Qué supuestos se necesitan para que el procedimiento usado sea válido?

c) Construir un intervalo de confianza para la diferencia de medias poblacionales.

Ejercicio 7.10

Para probar el efecto de distintas pasturas en el aumento de peso de novillos Aberdeen Angus, se seleccionaron 70 animales. 35 de ellos fueron elegidos al azar y se los alimentó durante 140 días con Triticale. Los otros 35 se alimentaron por igual período con Mijo. El promedio de aumento diario de peso en kg. fue de 0.65 con una desviación estándar de 0.08 kg. para el primer grupo y de 0.80 kg. con una desviación de 0.10 kg. para el segundo.

¿Existen diferencias significativas en el aumento de peso producido por estas dietas, trabajando con un nivel de significación del 1%?

Ejercicio 7.11

Para probar la eficacia de un tratamiento de poda en un bosque de Raulí, un investigador decide comparar el incremento del diámetro de los fustes de los árboles podados, con el incremento en árboles sin poda. Para ello se localizan 20 lotes de los cuales a 10 se los poda y al resto no. Al cabo de 3 años se obtienen los incrementos promedio para cada lote siendo los resultados los siguientes (en cm):

Stand con poda:	0.29	0.305	0.28	0.32	0.35	0.297	0.30	0.298	0.315	0.324
Stand sin poda:	0.30	0.303	0.27	0.30	0.32	0.31	0.28	0.302	0.298	0.301

¿Cuál es el efecto de la poda? Trabaje con un nivel de significación del 5%.

Ejercicio 7.12

A los fines de determinar los efectos de la restricción alimentaria en la química sanguínea de vacunos se midieron los metabolitos Calcio (Ca) y Fósforo (P) en sangre. El experimento se realizó tomando un lote de novillos de 180 kg. de peso promedio. De ellos, se eligieron aleatoriamente 10 para constituir el lote control (no restringidos) que eran alimentados con centeno a voluntad. El otro lote (restringidos) se conformó por los 10 animales restantes. La restricción consistió de dejar los novillos en pastoreo por 3 horas/día y luego pasarlos a corral pelado.

Se obtuvieron los siguientes resultados:

CALCIO		FOSFORO	
Restringidos	No restringidos	Restringidos	No Restringidos
6.93	5.99	7.24	8.69
8.42	8.82	7.46	6.13
8.55	8.82	7.59	6.79
8.69	8.82	7.73	6.79
8.82	8.95	7.86	6.93
8.82	8.95	8.26	7.59
8.95	9.05	8.39	7.86
8.95	9.34	8.39	9.06
9.61	9.34	8.53	9.59
9.10	10.66	8.53	9.73

a) ¿Cuál es la prueba apropiada para evaluar el efecto de la restricción en cada metabolito?

b) Probar los supuestos necesarios para la prueba anterior

c) ¿Altera la restricción alimentaria los parámetros sanguíneos? Utilizar $\alpha = 0.05$.

Ejercicio 7.13

La siguiente tabla presenta los resultados de una experiencia conducida para probar la hipótesis de que una dieta rica en lecitina favorece la producción de leche, en vacas de la raza Holando-Argentino. En este experimento se seleccionaron 18 tambos homogéneos en cuanto al manejo, de los cuales 9 fueron asignados aleatoriamente para recibir un suplemento de lecitina y los restantes actuaron como control. Debido a fallas en el seguimiento de uno de los tambos que no recibía el suplemento de lecitina, sus datos fueron descartados.

Los resultados, expresados en lts/día promedio por vaca son los siguientes:

Sin Lecitina	13.0	14.5	16.0	15.0	14.5	15.2	14.1	13.3	
Con Lecitina	17.0	16.5	18.0	17.3	18.1	16.7	19.0	18.3	18.5

Sean μ_{SL} la media de producción diaria de leche para animales de la raza Holando Argentino alimentados normalmente y μ_{CL} la media de producción de los animales alimentados con una dieta rica en lecitina. En base a los datos experimentales verificar la hipótesis: H_0: $\mu_{CL} = \mu_{SL}$ vs. H_1: $\mu_{CL} \geq \mu_{SL}$ (utilice $\alpha = 0.05$)

¿Cómo se informa el resultado de este ensayo?

Ejercicio 7.14

Un investigador supone que el estrés que se produce en vacas fistuladas puede disminuir los niveles de fósforo en sangre. Para probar su hipótesis selecciona 8 vacas y a cada una de ellas le extrae una muestra de sangre antes de la fistulación y otra muestra después. Los resultados son:

Vaca	1	2	3	4	5	6	7	8
Antes de la fistulación.	8.69	7.13	7.79	7.93	7.59	7.86	9.06	9.59
Después de la fistulación	7.24	7.10	7.80	7.95	7.50	7.79	9.00	9.48

¿Qué conclusión se puede extraer acerca de la fistulación? Utilizar $\alpha = 0.01$.

Ejercicio 7.15

Un criadero de semillas interesado en evaluar el comportamiento bajo riego de 2 híbridos de maíz realizó el siguiente ensayo: se tomaron 2 surcos de 50 m. y se delimitaron 10 sectores de 5 m. cada uno. Se sabe que el perfil de infiltración del agua es distinto a lo largo del surco de riego. Para evitar que este factor afecte la evaluación del rendimiento de los híbridos, en cada uno de los sectores de 5 metros de surco se asignaron aleatoriamente cada uno de ellos. Los datos obtenidos en qq/ha fueron:

Sector	1	2	3	4	5	6	7	8	9	10
Híbrido 1	123	121	119	115	111	105	106	114	120	127
Híbrido 2	127	130	118	117	114	110	115	120	125	133

Concluir acerca del comportamiento de los híbridos bajo riego. Utilizar $\alpha = 0.05$.

8

Análisis de la Varianza

Introducción

El Análisis de la Varianza -ANAVA- es, probablemente, la herramienta de inferencia estadística más utilizada en las investigaciones científico-técnicas en el campo de las ciencias biológicas en general y en las agropecuarias en particular. El ANAVA es un método estadístico cuya finalidad es probar hipótesis referidas a los parámetros de posición de dos o más poblaciones en estudio.

Definiciones preliminares

A continuación se dan un conjunto de definiciones necesarias para el tratamiento del tema.

Definición 8.1: Unidad experimental

Se llama **unidad o parcela experimental** a la mínima porción del material experimental sobre el cual un tratamiento puede ser realizado.

Por ejemplo, en un ensayo comparativo de rendimientos de trigo donde se desean evaluar 3 variedades se puede disponer de 30 parcelas de 1 m^2 cada una (unidades experimentales). Al final de la experiencia las plantas de cada parcela se cosecharán y en base a ello se realizará una medición del rendimiento en cada unidad.

Es importante conducir las experiencias de forma tal que las unidades experimentales generen información independiente. Así por ejemplo, para que el rendimiento de cada parcela sea independiente del rendimiento en las parcelas vecinas, se recurre a la aleatorización de las variedades a las parcelas. Además, es común en la investigación agropecuaria dejar espacio suficiente entre una parcela y otra para evitar dependencias o no dejar espacios libres, con el fin de simular mejor las condiciones reales de cultivo, y luego evaluar sólo el sector central de cada parcela. Esta técnica se conoce con el nombre de "bordura".

Definición 8.2: Tratamiento

Se denomina **tratamiento** al conjunto de acciones que se aplican a las unidades experimentales con la finalidad de observar como responden a éstas.

En la definición dada de tratamiento se dice que son acciones que "se aplican" a las unidades experimentales.... pero, ¿de qué forma se establece cuál unidad experimental va a recibir tal o cuál tratamiento? El procedimiento usual es asignar aleatoriamente los tratamientos a las unidades experimentales.

Definición 8.3: Variable aleatoria observada o respuesta

Se llama **variable aleatoria observada o respuesta** a la medida u observación que se obtiene de cada una de las unidades experimentales.

Retomando el ejemplo anterior, los tratamientos consisten en sembrar tres variedades de trigo en las parcelas experimentales y observar la respuesta: rendimiento de la parcela. Se dice en este caso que el *factor* tratamiento (variedad) tiene 3 *niveles.*

Las observaciones reales bajo cada tratamiento se asocian teóricamente a una distribución subyacente, así, si hay *a* tratamientos en estudio se tendrán *a* distribuciones. El conjunto de unidades experimentales que reciben un mismo tratamiento se asimila a una muestra aleatoria simple (m.a.s.) desde la distribución subyacente, ya que la variable aleatoria observada en cada unidad experimental es teóricamente independiente de la registrada en las otras.

Definición 8.4: Repetición

Se llama **repetición** a cada realización de un tratamiento

Prosiguiendo con el ejemplo, si se asignan 10 parcelas a cada cultivar, se tendrán **10 repeticiones** para cada tratamiento. Si además del factor cultivar se quisiera probar como afecta al rendimiento la aplicación de tres dosis de un mismo fertilizante, se estará en presencia de otro factor con tres niveles. Multiplicando el número de niveles del *factor cultivar* por el número de niveles del *factor dosis* del fertilizante se tienen un total de nueve tratamientos. Estos experimentos, donde los tratamientos son definidos a partir de la combinación de factores, se conocen como *experimentos factoriales*, en este caso bifactorial. Experiencias similares con más de dos factores se denominan experiencias multifactoriales y no son objeto de estudio en este libro.

La técnica de análisis de la varianza presupone un modelo para la variable respuesta. Este modelo recibe el nombre genérico de **modelo lineal**. A continuación se presenta su definición y se explican cada uno de los términos y sus propiedades estadísticas.

Definición 8.5: Modelo lineal

Se denomina **modelo lineal** de ANAVA (a una vía de clasificación) para la observación Y_{ij} a:

$$Y_{ij} = \mu + \tau_i + \varepsilon_{ij}\,, \quad \text{con } i=1,...,a \text{ y } j=1,..,n$$

donde:

Y_{ij} es la j-ésima observación del i-ésimo tratamiento

μ es la media general de las observaciones

τ_i es el efecto del i-ésimo tratamiento

ε_{ij} es una variable aleatoria normal independientemente distribuida con esperanza 0 y varianza $\sigma^2\ \forall i,j$.

En la Figura 8.1 se esquematizan *a*=3 distribuciones centradas en sus esperanzas, denotadas por μ_i, y se representan parámetros del modelo lineal.

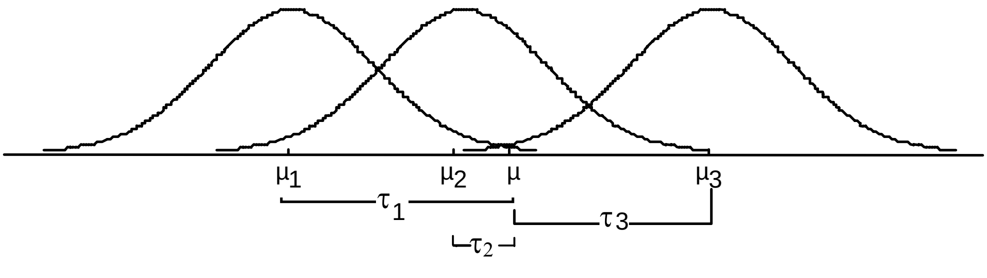


Figura 8.1: Representación de 3 funciones de densidad, mostrando el punto de equilibrio de todas ellas (μ), las esperanzas de cada una de ellas (μ_i) y los corrimientos de las esperanzas respecto del punto de equilibrio representando o efectos de tratamiento (τ_i).

La media general (μ) es el centro de equilibrio de todas las distribuciones y se trata de un *parámetro fijo*. El efecto del tratamiento (τ_i) se presenta como un corrimiento respecto de la media general y en el modelo conocido como de ANAVA de efectos fijos se asume constante. El efecto del tratamiento 1 (τ_1) es la diferencia que hay entre la media del tratamiento 1 y la media general. La hipótesis nula del ANAVA postula la igualdad de medias de todos los tratamientos comparados. Si la hipótesis nula del ANAVA fuera verdadera las *a* distribuciones estarían centradas sobre la misma esperanza, es decir, en μ. Los valores de la variable aleatoria ε_{ij} representan las diferencias entre observaciones individuales y las esperanzas de la distribución de la cual proviene la observación.

El modelo lineal presentado corresponde a un diseño completamente aleatorizado a un criterio de clasificación. En el Capítulo 10 se presentarán otros modelos que incluyen más *parámetros* para denotar la mayor complejidad estructural del diseño experimental. Esto implica que no existe un único modelo lineal y la selección de un modelo para cada problema forma parte del arte del análisis de datos experimentales. Si el modelo propuesto no es adecuado se parte de una muy mala base para probar las hipótesis planteadas.

Existen dos tipos básicos de modelos lineales de ANOVA a un criterio de clasificación: de efectos fijos y aleatorios, dependiendo de la naturaleza aleatoria o no de los efectos de tratamiento. En esta obra sólo se consideran los modelos de efectos fijos.

El análisis de la varianza de efectos fijos a un factor de clasificación

El objetivo del ANAVA de efectos fijos es contrastar la hipótesis de que los efectos de tratamientos son nulos versus que al menos uno no lo es. En términos estadísticos:

H_0: $\tau_1=...=\tau_a= 0$ vs. H_1: Al menos un tratamiento tiene efecto no nulo.

Otra forma de enunciar estas hipótesis es *que las medias de los tratamientos que se comparan son idénticas vs. que no lo son*. La técnica de ANAVA es sensible a las propiedades estadísticas de los errores del modelo lineal y supone que los datos observados son independientes unos de otros y que las observaciones bajo cada tratamiento tienen distribución normal centrada en su esperanza $(\mu+\tau_i)$ y varianza σ^2, idéntica para toda observación (*homogeneidad de varianzas*). El no cumplimiento de estas propiedades, conocidas como *supuestos*, puede invalidar la inferencia que se pueda realizar a partir de esta técnica.

Fundamentos del análisis de la varianza de efectos fijos

Si se toma una muestra aleatoria simple de cada una de *a* distribuciones con idéntica varianza, entonces las *a* varianzas muestrales estiman al mismo parámetro y el promedio ponderado de estas varianzas es un buen estimador de σ^2. Por otra parte, si además de idénticas varianzas se pide idénticas esperanzas (hipótesis nula en el ANAVA), las *a* medias muestrales son estimaciones de la misma media poblacional y tienen varianza $\sigma^2_{\bar{X}} = \frac{\sigma^2}{n}$. Bajo estas condiciones, a partir de la varianza de las medias muestrales se puede obtener otra estimación de σ^2. En consecuencia, si el *supuesto de idéntica varianza* y la *hipótesis de igualdad de medias* son ciertos, se tienen, a partir de una muestra, dos estimadores independientes de la varianza poblacional. Si por el contrario, la *hipótesis de igualdad de medias* no es cierta, entonces la varianza estimada a partir de las medias incluirá una fuente de variación debida a la diferencia de los parámetros de posición de las distribuciones muestreadas. Luego, la comparación del promedio ponderado de las varianzas muestrales con el estimador obtenido a partir de la varianza de las medias muestrales es la clave del método de análisis de la varianza y de allí su nombre. Cuando la *hipótesis de igualdad de medias* falla, el estimador obtenido a partir de la varianza de las medias muestrales es más grande que lo esperado y en consecuencia sirve para detectar la desigualdad de las esperanzas de las distribuciones que se comparan.

Cuadrados medios y prueba de hipótesis

Definición 8.6: Cuadrado Medio Dentro o del Error

Si X_{i1}, X_{i2}, ...X_{in} es muestra obtenida bajo el tratamiento i-ésimo y se tienen muestras para *a* tratamientos, entonces, si σ^2 representa la varianza de la distribución bajo cualquier tratamiento, se llamará **Cuadrado Medio Dentro** (CMD) al promedio ponderado de las *a* varianzas estimadas en cada tratamiento

$$CMD = \frac{(n_1 - 1)S_1^2 \; + \; ... \; + \; (n_a - 1)S_a^2}{(n_1 - 1) \; + \; ... \; + \; (n_a - 1)}$$

Se puede probar que el Cuadrado Medio Dentro es un estimador insesgado de σ^2, es decir E(CMD) = σ^2.

Nota: El nombre Cuadrado Medio Dentro proviene del hecho que es un promedio de magnitudes cuadráticas. Este ofrece una medida de la variabilidad promedio que hay dentro de cada tratamiento y mide la variabilidad de unidades experimentales tratadas de la misma forma (error experimental), por ello también se suele llamar cuadrado medio del error.

Definición 8.7: Cuadrado Medio Entre o Cuadrado Medio de Tratamiento.

Si X_{i1}, X_{i2}, ...X_{in} es una muestra obtenida bajo el tratamiento i-ésimo y se tienen muestras para cada uno de *a* tratamientos, es posible obtener la varianza de las medias muestrales $S_{\bar{X}}^2$ y a partir de ésta, encontrar un estimador de σ^2 que se denomina **Cuadrado Medio Entre o Cuadrado Medio de Tratamiento** (CME)

$$CME = S_{\bar{X}}^2 \cdot n .$$

A diferencia del CMD que es un estimador incondicional de σ^2, el CME estima a σ^2 *sólo si las esperanza de los tratamientos que se comparan son iguales* (H_0 verdadera) de lo contrario estima a $\sigma^2 + c\sum_{i=1}^{a} \tau_i^2$, con *c* una constante mayor que 0.

Luego, CME es un estimador insesgado de σ^2 sólo si H_0 es verdadera, de lo contrario estima a σ^2 más una cantidad que representa una medida de la magnitud de los efectos de tratamiento.

Si H_0 es verdadera todo τ_i=0 y por lo tanto la componente añadida por los efectos de tratamiento se anula y la esperanza del CME es σ^2. Luego el CMD y el CME son estimadores independientes de σ^2 bajo H_0[8].

¿Cómo establecer si la hipótesis nula de igualdad de efectos de tratamientos es verdadera o falsa? La respuesta está al alcance del lector si piensa sobre el estadístico F utilizado en la prueba de hipótesis de igualdad de dos varianzas.

Sean σ_E^2 y σ_D^2 las varianzas estimadas por el CME y el CMD respectivamente, luego bajo la hipótesis de igualdad de medias de tratamiento, $\sigma_E^2 = \sigma_D^2$ de lo contrario $\sigma_E^2 > \sigma_D^2$ por lo tanto las hipótesis de una prueba estadística son las siguientes:

$$H_0 : \sigma_E^2 = \sigma_D^2 \quad \text{vs} \quad H_1 : \sigma_E^2 > \sigma_D^2$$

La prueba consiste en calcular el estadístico F utilizando los estimadores de σ_E^2 y σ_D^2 de la siguiente forma:

$$F = \frac{CME}{CMD}$$

Este estadístico tiene, bajo H_0, una distribución $F_{(a-1),(N-a))}$ con $N = \sum_{i=1}^{a} n_i$. Luego, para un

[8] La demostración de que estas estimaciones son independientes está fuera del perfil de este libro.

nivel de significación α, si F es mayor que el cuantil (1-α) de la distribución $F_{(a-1),(N-a)}$ se rechaza H_0, implicando que H_1 es verdadera. El rechazo de H_0 implica que $\sum_{i=1}^{a}\tau_i^2$ es distinto de 0 y por lo tanto, que algún $\tau_i \neq 0$; luego se concluye que no todas las medias de tratamiento son iguales o que al menos un tratamiento tiene efecto distinto de 0.

En síntesis, el ANAVA se basa en dos estimadores independientes de la varianza de las observaciones: uno basado en la variabilidad dentro de los tratamientos, y otro basado en la variabilidad entre los tratamientos. Si no hay diferencias entre las medias de los tratamientos, estos dos estimadores estiman al mismo parámetro, de lo contrario el segundo tiende a ser mayor cuanto mayor es la diferencia entre los tratamientos. Luego, a pesar de que la hipótesis de interés del ANAVA se refiera a la igualdad de las esperanzas de dos o más distribuciones, la técnica del ANAVA se basa en la comparación de varianzas para inferir acerca de la igualdad de las esperanzas.

La partición de la suma de cuadrados y la tabla del ANAVA

A fin de presentar el procedimiento para el análisis de la varianza se introduce la notación que describe los datos. Supóngase que se tienen *a* tratamientos, que la variable de respuesta se representa con la letra "Y", que se dispone de *n* repeticiones para cada tratamiento, y que la asignación de los mismos a las unidades experimentales se realiza bajo un diseño completamente aleatorizado. Bajo estas condiciones los datos pueden representarse según la Tabla 8.1.

Tabla 8.1: Estructura típica de una tabla de datos de un experimento unifactorial con diseño completamente aleatorizado.

Tratamientos					Media
1	y_{11}	y_{12}	...	y_{1n}	$\bar{y}_1$
2	y_{21}	y_{22}	...	y_{2n}	$\bar{y}_2$
:	:	:	:::	:	
a	y_{a1}	y_{a2}	...	y_{an}	$\bar{y}_n$

Cada dato está representado por y_{ij}, y hace referencia a la observación j-ésima tomada bajo el tratamiento i-ésimo. Por ejemplo, el dato y_{12} representa a una observación realizada sobre la unidad experimental número 2 del tratamiento designado como 1; con y_{an} a la observación de la unidad experimental n-ésima del tratamiento a-ésimo en general.

El análisis de la varianza se presenta en una Tabla conocida como Tabla de Análisis de la Varianza en la que se resumen los estadísticos y cálculos básicos para obtener el CME y el CMD, estadísticos claves para la prueba de hipótesis. En la Tabla 8.2, $N = \sum_{i=1}^{a} n_i$ y la notación $y_{i\bullet}$ indica sumar sobre el índice reemplazado por el punto, esto es: $y_{i\bullet} = \sum_{j=1}^{n_i} y_{ij}$.

Tabla 8.2: Fórmulas de trabajo para el análisis de la varianza de un experimento unifactorial con diseño completamente aleatorizado.

Fuente de Variación	Suma de Cuadrados	Grados de Libertad	Cuadrado Medio	F Obs.
Entre Tratamientos	$SCE=\sum_{i=1}^{a} \frac{(y_{i\bullet})^2}{n_i} - \frac{(y_{\bullet\bullet})^2}{N}$	gle= a -1	$CME=\frac{SCE}{gle}$	$\frac{CME}{CMD}$
Dentro (Error Experimental)	SCD=SCT-SCE	gld= N - a	$CMD=\frac{SCD}{gld}$	
Total	$SCT=\sum_{i=1}^{a} \sum_{j=1}^{n_i} y_{ij}^2 - \frac{(y_{\bullet\bullet})^2}{N}$	glt= N - 1		

En la columna titulada "Fuentes de Variación" se destacan tres celdas con sus correspondientes títulos. En ellas se indican los contenidos de las celdas dentro de la fila respectiva. En la fila titulada "Entre Tratamientos" existen cuatro celdas, en las que se calculan las siguientes cantidades: *Suma de Cuadrados Entre Tratamientos (SCE)*, *Grados de Libertad de la suma de cuadrados entre tratamientos (gle)*, *Cuadrados Medios Entre Tratamientos (CME)* y el estadístico F correspondiente al cociente del CME/CMD. La fila titulada "Dentro (Error Experimental)" se completa con las siguientes cantidades: *Suma de Cuadrados Dentro de Tratamientos (SCD)*, *Grados de Libertad de la suma de cuadrados dentro de tratamientos (gld)* y *Cuadrado Medio Dentro de Tratamientos(CMD)*. En la titulada "Total" se completa con la *Suma de Cuadrados Total* (SCT) y *Grados de Libertad Totales (glt)*.

Ejemplo 8.1

El porcentaje de humedad relativa (HR) es determinante para el ataque de hongos en semillas. Para evaluar la susceptibilidad de las semillas maní al ataque de un hongo se realizó un ensayo en cámaras de cría con tres porcentajes de HR: 70%, 80% y 90%. Cinco observaciones fueron tomadas para cada porcentaje de HR, registrándose el número de semillas atacadas en un grupo de 100 semillas (unidad experimental). Las observaciones se presentan en la Tabla 8.3.

Tabla 8.3: Datos obtenidos de un ensayo sobre el efecto de hongos en la semilla de maní.

Porcentaje de HR	Observaciones (Número de semillas atacadas)					Totales de Tratamiento $y_{i\bullet}$
70	7	6	9	5	9	36
80	12	15	17	18	20	82
90	14	16	18	21	15	84
						$y_{\bullet\bullet} = 202$

Los cálculos preliminares del ANAVA son:

$$SC_{Total} = 7^2 + 6^2 + 9^2 + ... + 21^2 + 15^2 - \frac{202^2}{15} = 375.73$$

$$SC_{Entre} = \frac{36^2+82^2+84^2}{5} - \frac{202^2}{15} = 294.93$$

$$SC_{Dentro} = 375.73 - 294.93 = 80.8$$

Así, la tabla de ANAVA correspondiente es:

Tabla 8.4: Análisis de Varianza para un ensayo sobre el efecto de hongos en la semilla de maní según tratamientos de humedad relativa.

Fuente de Variación	Suma de Cuadrados	Grados de Libertad	Cuadrados Medios	F
Entre Tratamientos	294.93	2	147.46	21.91
Dentro (Error Experimental)	80.8	12	6.73	
Total	375.73	14		

Si $\alpha = 0.05$, luego el punto crítico que delimita la zona de aceptación y rechazo de H_0 es $F_{(2,12;\ 0.95)} = 3.88$. Como F= 21.91> $F_{crítica}$ se concluye, con un nivel de significación del 5%, que se rechaza la hipótesis nula de igualdad de medias de número de semillas atacadas para los 3 porcentajes de HR, por lo tanto al menos una de las HR produce un grado de ataque de hongos diferente de los restantes.

Pruebas "a posteriori"

Si se rechaza la hipótesis nula del ANAVA, la pregunta que sigue es ¿cuál o cuáles de las medias poblacionales en estudio son las diferentes?

Si el número de tratamientos es suficientemente grande, es probable que la diferencia entre la media mayor y la menor sea declarada como significativa por una prueba T aún cuando la H_0 no fue rechazada en el ANAVA. Así, realizando comparaciones de a pares usando la prueba T, cada una con un nivel α, la probabilidad de rechazar incorrectamente H_0, al menos una vez, incrementaría con el número de tratamientos. Luego, teniendo como objetivo controlar α, varios procedimientos de comparaciones múltiples 'a posteriori' han sido propuestos en la literatura desde la década del '50.

Existe una gama muy amplia de alternativas para llevar adelante este tipo de pruebas, entre las que se destacan la de Tukey (Tukey, 1949), la de Scheffé (Scheffé, 1953), la de Duncan (Duncan, 1955), la de Dunnet (Dunnet, 1964), y la de Fisher (Fisher, 1966), entre otras. Se darán a continuación las pruebas de Tukey y de Fisher. Estas pruebas no agotan las múltiples posibilidades de elección de métodos de comparaciones, pero representan un método conservador, es decir que controla la tasa de error tipo I (Tukey), y uno que no lo es tanto (Fisher).

El test de Tukey

El test de Tukey examina con un mismo estadístico todas las diferencias de medias muestrales en estudio. Si hay *a* medias, luego habrá $\binom{a}{2} = \frac{a!}{(a-2)!\,2!}$ diferencias de medias posibles.

El estadístico de Tukey es el siguiente:

$$\text{DMSt} = q_{a,gld;\,(1-\alpha)}\sqrt{\frac{\text{CMD}}{n}}$$

donde $q_{a,gld;\,(1-\alpha)}$ es el cuantil (1-α) que se obtiene de la distribución de Rangos Studentizados (ver Anexo) para *a* tratamientos y los grados de libertad dentro; α es el nivel de significación en base al cual se rechazó la H_0 del ANAVA y *n* es el número de repeticiones en base a las que se calculan las medias muestrales. Si el tamaño de muestra no fuera el mismo para cada tratamiento, deberá reemplazarse *n* por la media armónica de los $\{n_i\}$, esto es:

$$n_0 = \frac{a}{\sum_{i=1}^{a}\frac{1}{n_i}}$$

Si el valor absoluto de la diferencia entre un par de medias supera a DMSt, se dice que esta diferencia es estadísticamente significativa. Se concluirá en consecuencia que las esperanzas asociadas a esa diferencia son distintas con un nivel de significación α.

Cabe destacar que cuando los tamaños muestrales son muy diferentes, el test de Tukey puede dejar de ser confiable, caso en el cual podría utilizarse algún procedimiento de contraste múltiple que considere tal situación, como el de Scheffé (1953).

Retomando el Ejemplo 8.1 recuérdese que se había concluido que los diferentes porcentajes HR producían un diferente grado de ataque del hongo sobre la semilla de maní. La pregunta que sigue es ¿cuál o cuáles de ellos producen ataques diferentes? Para dar respuesta a ello se utilizará la prueba de Tukey. Aunque no es necesario, se puede construir una matriz de valores absolutos de las diferencias entre medias como la que se muestra a continuación.

Tabla 8.5: Matriz de diferencias de medias del Ejemplo 8.1.

	$\bar{x}_1$	$\bar{x}_2$	$\bar{x}_3$
$\bar{x}_1$		9.2	9.6
$\bar{x}_2$			0.4
$\bar{x}_3$			

El segundo paso consiste en calcular el estadístico de Tukey. Para el ejemplo, $a = 3$, gld=12 y α=0.05 (el mismo usado en el ANAVA), $q_{a,gld;(1-\alpha)} = 3.77$; CMD = 6.73 (Tabla 8.4) y $n = 5$ (número de repeticiones). Así se tiene:

$$\text{DMSt} = 3.77\sqrt{\frac{6.73}{5}} = 4.37$$

Para terminar con esta prueba basta controlar qué diferencias entre medias muestrales son mayores que 4.37 para concluir que las esperanzas que estiman difieren entre sí con un nivel de significación del 5%. Revisando la matriz de diferencias de medias se puede verificar que :

$\mu_1 \neq \mu_2$ por cuanto $|\bar{x}_1 - \bar{x}_2| = 9.2 > 4.37$;

$\mu_1 \neq \mu_3$ por cuanto $|\bar{x}_1 - \bar{x}_3| = 9.6 > 4.37$;

$\mu_2 = \mu_3$ por cuanto $|\bar{x}_1 - \bar{x}_3| = 0.4 < 4.37$;

Luego, el grado de ataque que se produce con un 80% de HR no difiere del que se produce con 90% de HR, mientras que con 70% de HR se produce un ataque significativamente menor que con 80 y 90%.

Sintetizando se podría afirmar con un 95% de confianza que el menor grado de ataque se produce con 70% de HR.

Prueba de Fisher

La prueba de Fisher es similar en su procedimiento a la prueba de Tukey, pero el estadístico de la prueba es diferente. En vez de usar los cuantiles de la distribución de rangos estudentizados utiliza los cuantiles de una t con los grados de libertad del cuadrado medio dentro de tratamientos y es particular para cada comparación de medias ya que depende del número de repeticiones por tratamiento. Luego, la diferencia mínima significativa entre el tratamiento i-ésimo y el tratamiento j-ésimo está dada por:

$$DMSf_{ij}=t_{gld;(1-\alpha/2)}\sqrt{CMD\ \frac{n_i+n_j}{n_i\ n_j}}$$

Para el Ejemplo 8.1 $t_{12;(0.975)}$= 2.179, CMD=6.73 y n_i=n_j=5 $\forall$ij, luego la diferencia mínima significativa por Fisher es para todas las comparaciones

$$DMSf_{ij}=2.179\sqrt{6.73\ \frac{5+5}{5.5}}\ =3.58$$

Es interesante mostrar que mientras para Fisher la diferencia mínima significativa es 3.58, para Tukey es 4.37. Esto implica que con Fisher es más fácil rechazar la hipótesis de igualdad de medias que con Tukey, por esta razón se dice que este último es más conservador (menor error tipo I) y el primero más potente (menor error tipo II).

Verificación de supuestos del análisis de la varianza

Como se recordará, los supuestos del análisis de la varianza se refieren a las propiedades estadísticas de los errores. Usualmente se suponen normales con esperanza cero, varianza común e independientes. La verificación de estas propiedades garantiza que las conclusiones del ANAVA estén acotadas en sus Errores Tipo I y Tipo II. Existen distintas técnicas de validación de supuestos, pero las que se presentan aquí se basan en los predictores de los errores, es decir los residuos.

Una vez calculados los predictores se puede verificar el cumplimiento de los supuestos de normalidad, independencia y homogeneidad de varianzas de los ε_{ij}, mediante pruebas de hipótesis e interpretaciones gráficas.

A continuación se da una definición de residuo para el modelo de análisis de la varianza.

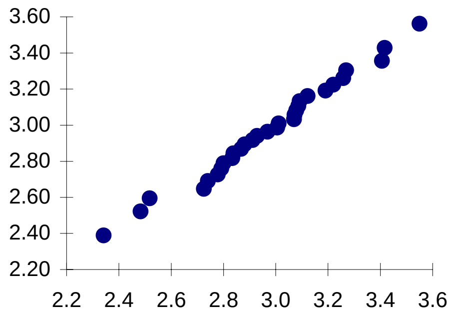
Definición 8.8: Residuo

Se llamará **residuo** de la observación j-ésima del tratamiento i-ésimo al predictor de ε_{ij}, que se denota por e_{ij} , y se obtiene como la diferencia entre el valor observado y el valor predicho por el modelo. En el modelo presentado: $e_{ij} = y_{ij} - \bar{y}_i$

Una vez calculados los residuos del experimento se pueden verificar los supuestos y evaluar si el modelo lineal es el correcto. Si no es este el caso, es decir, si se detecta falta de independencia o de normalidad o de homogeneidad de varianzas, el modelo elegido no es adecuado para el análisis. A continuación se considera cada uno de los supuestos, y cómo evaluarlos mediante interpretación gráfica.

Normalidad: tomando los residuos como datos, una de las técnicas más usadas es construir un Q-Q plot normal. Mediante esta técnica (ver Capítulo 1) se obtiene un diagrama de dispersión en el que, si los residuos son normales y no hay otros defectos del modelo, entonces se alinean sobre una recta a 45° como se muestra en la siguiente figura. La presencia de ligeras violaciones de este supuesto no es muy grave, no afectándose de forma importante la probabilidad de cometer Error de Tipo I, pero en algunos casos puede elevarse demasiado la probabilidad de cometer error Tipo II. La Figura 8.2 ilustra un Q-Q plot de residuos obtenidos a partir de un modelo con errores normales homocedásticos.

Figura 8.2: Q-Q plot (normal)



Independencia: Una ayuda valiosa para estudiar la falta de independencia entre los errores es realizar un gráfico de los residuos según la secuencia en el tiempo o espacio físico en que han sido colectados los datos. Si los residuos aparecen en secuencias de varios valores positivos seguidos de varios valores negativos puede ser un indicio claro de la falta de independencia. Otro posible patrón indicativo de falta de independencia es una sucesión alternante de residuos positivos y negativos. Siempre que se detecte cualquier patrón en este gráfico se debe sospechar del incumplimiento del supuesto de independencia. Un ejemplo de un gráfico en el que no se observan patrones sospechosos en la distribución de residuos es el siguiente.

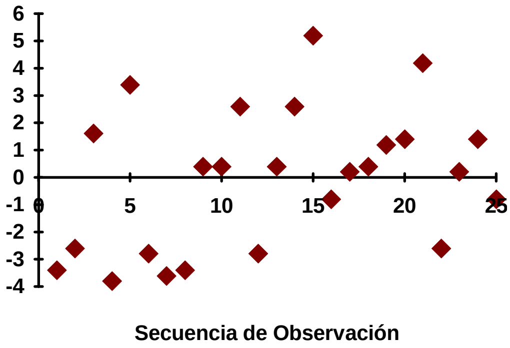
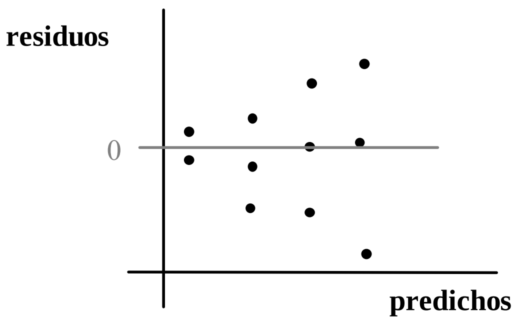


Figura 8.3: Dispersión de los residuos en función de la secuencia de observación.

La falta de independencia es un problema potencialmente peligroso y difícil de corregir, por lo que es importante prevenirlo. La aleatorización en la asignación de los tratamientos a las unidades experimentales, en la secuencia de medición de los resultados del ensayo, o en cualquier otra etapa experimental que pueda introducir una fuente sustancial de error, es uno de los métodos más eficaces de controlar la falta de independencia.

Homogeneidad de varianzas: haciendo un gráfico de dispersión de residuos vs. predichos se debe observar una nube de puntos sin patrón alguno para comprobar que las varianzas son homogéneas. Un patrón típico que indica falta de homogeneidad en las varianzas, se muestra en la Figura 8.4, ya que a medida que crecen los valores predichos por el modelo, aumenta la dispersión de los residuos.

Figura 8.4 Gráfico de Residuos en función de Predichos en un ejemplo con falta de homogeneidad de varianzas



Nota: Se debe ser cuidadoso en la interpretación de estos gráficos ya que el patrón mostrado por la Figura 8.4 se puede presentar cuando los tamaños de muestras son distintos en cada tratamiento, no indicando necesariamente heterogeneidad de las varianzas.

Ejercicios

Ejercicio 8.1

Se desea conocer el efecto de las cepas de inoculantes sobre el contenido de nitrógeno de plantas de trébol rojo. Para ello se dispone de 30 macetas de trébol rojo en un invernadero. Se asignan al azar 5 macetas para cada una de las cepas y se procede a inocularlas. Los resultados son los siguientes (en mg. de nitrógeno):

Cepa I	Cepa II	Cepa III	Cepa IV	Cepa V	Cepa VI
19.4	17.7	09.1	18.6	11.6	16.9
27.0	24.3	11.9	18.8	11.8	17.3
32.1	24.8	15.8	20.5	14.2	19.1
32.6	25.2	17.0	20.7	14.3	19.4
33.0	27.9	19.4	21.0	14.4	20.8

a) Plantear H_0 y H_1

b) Realizar el Análisis de la Varianza ($\alpha = 0.05$)

c) Si corresponde, realizar una prueba "a posteriori".

Ejercicio 8.2

En un estudio sobre el efecto de la adición de azúcares sobre diámetro de secciones de poroto criados en un medio de cultivo, se obtuvieron los siguientes datos:

Control	75	67	70	75	65	71	67	67	76	68
Glucosa	57	58	60	59	62	60	60	57	59	61
Fructosa:	58	61	56	58	57	56	61	60	57	58
Gluc. + Fruc.	58	59	58	61	57	56	58	57	57	59
Sacarosa	62	66	65	63	64	62	65	65	62	67

¿Qué se puede decir sobre el efecto de los distintos medios de cultivo? Concluir trabajando con un nivel de significación de 0.05.

Ejercicio 8.3

Se desea estudiar el efecto de la carga animal sobre la producción de materia seca en una pastura implantada. Para ello se divide un lote en 28 potreros y se asignan aleatoriamente 7 potreros a cada una de las 4 cargas animales en estudio (2 nov./ha., 4 nov./ha, 6 nov./ha. y 8 nov./ha.)

Los resultados fueron los siguientes expresados en toneladas de materia seca por hectárea.

								Media
carga 2	2.6	1.9	3.1	2.8	2.2	2.0	2.7	2.47
carga 4	3.3	3.6	3.0	3.5	3.2	3.9	3.4	3.41
carga 6	3.1	2.0	2.5	3.1	2.3	3.0	2.2	2.60
carga 8	2.5	2.3	2.8	1.8	2.7	2.6	2.0	2.39

a) Plantear un modelo lineal que permita recomendar alguna carga en especial.

b) ¿Qué supuestos se requieren para el análisis de este ensayo?

c) Realizar el análisis y concluya. Trabajar con un nivel de significación de 0.05.

Ejercicio 8.4

Se supone que buena parte de las diferencias entre las variedades A y B de una especie vegetal, se deben no a causas genéticas sino al efecto del medio ambiente donde se desarrollan. Para probar (parcialmente) esta hipótesis se realizó un experimento en el cual 10 lotes de cada variedad se hicieron crecer en un mismo ambiente. La *altura de planta* fue la variable que se registró y los datos son los siguientes:

										$\sum_i x_i$	$\sum_i x_i^2$	nj
Variedad A	15	12	8	14	16	16	9	15	11	14	130	1764	10
Variedad B	12	9	13	10	8	12	13	14	9	10	110	1248	10

a) Identificar las H_0 y H_1 y el modelo a adoptar.

b) Realizar una prueba T y un análisis de varianza, usando un nivel de significación del 5%. Comprobar que el valor de T^2 reproduce el valor del estadístico F.

c) ¿Qué se concluye sobre las diferencias varietales?

Ejercicio 8.5

Una empresa agrícola necesita establecer si le conviene fertilizar sus cultivos de soja y si es así, seleccionar el mejor fertilizante. Para este propósito se realizó un ensayo en un lote de 5 has., dividido en parcelas de 1/4 ha. cada una, asignando los tratamientos en forma aleatoria. Los rendimientos obtenidos (qq/ha) fueron:

Control (sin fertilizar)	Fert. A	Fert. B	Fert. C
23	30	28	27
20	32	36	25
22	29	31	24
20	35	32	28
21	33	34	26

a) Hacer una representación gráfica comparativa de los rendimientos

b) ¿Se recomendaría la fertilización?

c) De ser así, ¿cuál de los fertilizantes se recomendaría?

Ejercicio 8.6

En un experimento para evaluar suplementos en las dietas de ovejas se escogieron 16 ovejas al azar, de un rebaño, separándolas aleatoriamente, en grupos de 4 animales. Las 4 primeras (primer grupo) se suplementaron con el producto A, otras 4 con el producto B, otras 4 con el producto C y las restantes se dejaron como testigo, sin suplemento. Las medias de aumento de peso por animal al cabo de 100 días, fueron (en libras):

A = 55 B = 57 C = 63 Testigo = 52

Se realizó un ANAVA para este experimento con los siguientes resultados:

SCT = 646; gle=3

Completar la siguiente Tabla del ANAVA:

Fuente de Variación	Suma de Cuadrados	Grados de Libertad	Cuadrados Medios	F Observada	F Tabla
Entre					
Dentro					
Total					

Ayuda: Revea la definición 8.7 para calcular el cuadrado medio de tratamientos.

a) Si se justifica, realizar una prueba a posteriori trabajando con $\alpha = 0.05$.

b) Informar los resultados del ensayo utilizando gráficos adecuados.

Ejercicio 8.7

En una experiencia realizada para determinar si los pesos (mg) de las hembras adultas de *Drosophila permisilis*, criadas a 24ºC, resultan afectados por la densidad a la que se crían las larvas, se pesaron 10 ejemplares adultos de cada medio, obteniéndose los siguientes resultados:

Densidad larval	Peso medio	Varianza de los pesos	n_i
1	1.356	0.032	10
3	1.356	0.018	10
5	1.284	0.017	10
6	1.252	0.011	10
10	0.989	0.017	10
20	0.664	0.020	10

Realizar un análisis de la varianza para saber si existen diferencia estadísticamente significativas entre los pesos atribuibles a las distintas densidades larvales. Trabajar con $\alpha = 0.05$.

9

Análisis de Regresión Lineal

Introducción

El objetivo de este capítulo es introducir el análisis simultáneo de dos variables y adquirir criterios para el uso de las técnicas de regresión y correlación.

Hasta el capítulo anterior se han introducido métodos estadísticos que se pueden utilizar cuando el interés es analizar el comportamiento de una sola variable, eventualmente, bajo distintas condiciones. Por ejemplo, el rendimiento o la altura de las plantas de un cultivo con o sin riego. Pero frecuentemente se presentan situaciones donde se observan dos o más variables sobre cada unidad experimental y el interés se centra en la forma en que estas variables se relacionan.

Algunos ejemplos de relaciones funcionales que pueden ser de interés en agronomía son: la relación entre el rendimiento de un cultivo y la densidad de siembra, la relación entre la cantidad de suplemento dado y el aumento de peso que éste produce en un lote de animales, las dosis de un insecticida y la mortalidad de los insectos tratados, etc. En cada uno de estos casos se pueden plantear los siguientes interrogantes:

¿Existe alguna relación entre las variables?

Si se conoce el comportamiento de una de ellas, ¿se puede predecir el comportamiento de la otra?

La estadística aplicada ofrece dos herramientas que permiten dar respuesta a dichas cuestiones: el *Análisis de Regresión* y el *Análisis de Correlación*.

El ***Análisis de Regresión*** estudia la relación funcional que existe entre dos o más variables. Identifica el **modelo o función** que liga a las variables, estima sus parámetros y, eventualmente, prueba hipótesis acerca de ellos. Una vez estimado el modelo es posible predecir el valor de la variable denominada **variable dependiente** en función de la o las otras **variable/s independiente/s** y dar una medida de la precisión con que esa estimación se ha hecho.

Dependiendo del objetivo del estudio, los valores o **niveles** de la/s variable/s independiente/s pueden ser arbitrariamente modificados por el experimentador, es decir el investigador puede fijar los niveles de la variable independiente para los cuales desea estudiar la respuesta de la variable dependiente. El modelo hallado puede ser usado para predecir el comportamiento de la variable dependiente para otros niveles de la variable independiente, que pertenezcan al **dominio** del estudio.

El ***Análisis de Correlación lineal*** estudia el grado y sentido de la asociación lineal que hay entre un conjunto de variables y, a diferencia del análisis de regresión, no se identifica ni se estima explícitamente un modelo funcional para las variables, este siempre se supone lineal. El interés principal es medir la asociación entre dos variables aleatorias cualesquiera, sin necesidad de distinguir variables dependientes e independientes. Por ejemplo, puede quererse evaluar la intensidad de la asociación entre la cantidad de espiguillas por espiga de trigo y la longitud de las espigas. Se ha establecido que cuanto mayor es la longitud de las espigas mayor es el número de espiguillas por espiga. Obsérvese que, en el ejemplo, no se habla de relación funcional, ni tampoco se insinúa que la longitud de la espiga aumenta porque aumenta el número de espiguillas o viceversa, sólo se enfatiza la forma en que se comporta una variable en relación a la otra y el interés está centrado en medir la intensidad de esta asociación.

En el análisis de correlación, ninguna de las variables puede ser fijada por el experimentador, ya que éste podría seleccionar niveles de las variables que no son frecuentes y esto podría conducir a una estimación errada del grado de correlación.

Los **gráficos de dispersión** son útiles en la etapa exploratoria, tanto en el análisis de regresión como en el de correlación. La representación gráfica de los datos es frecuentemente el punto de partida de cualquier análisis que involucra más de una variable. En los gráficos de dispersión lo que se ve es una nube de puntos, donde cada punto representa una observación. La Figura 9.1 muestra los gráficos de dispersión usados en estudios de asociación entre dos variables donde además se ha dibujado sobre la nube de puntos, la posible función de ajuste de esos datos, es decir, se ha *identificado* el modelo funcional de la relación.

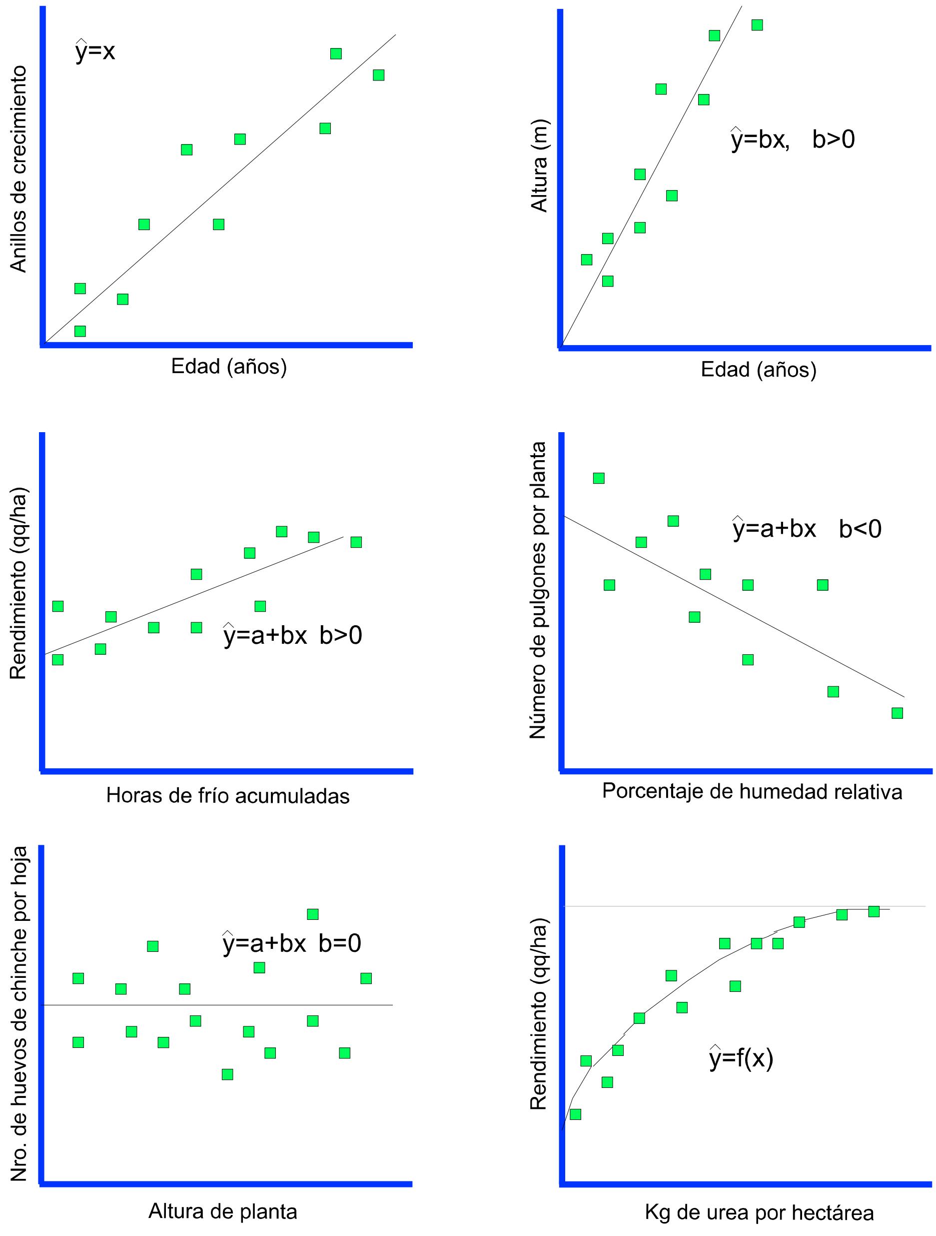


Figura 9.1: Gráficos de dispersión para diferentes modelos de relación entre dos variables.

Análisis de regresión lineal

El término "regresión" surgió de estudios de la herencia biológica realizados por Galton durante el siglo pasado. En su conocida experiencia, Galton notó que los padres altos tenían hijos cuya altura era mayor a la altura promedio, pero no eran más altos que sus padres. También, padres bajos tenían hijos con altura menor a la altura promedio pero eran más altos que sus padres. Esta tendencia de las características de los grupos de moverse, en la siguiente generación, hacia el promedio de la población o de **regresión hacia la media** fue descubierta por Galton. El término no tiene hoy el mismo significado que le dio Galton, pero se usa extensamente para referirse al estudio de relaciones funcionales entre variables cuando hay una componente aleatoria involucrada.

Al estudiar la relación entre dos o más variables surge la idea de encontrar una expresión matemática que la describa. Para el caso de dos variables, si se denota como Y a la variable que se supone **dependiente** y como X a la variable que se postula como **independiente,** resulta familiar utilizar el concepto de función y decir "*Y es función de X*", para indicar que de acuerdo a los valores asignados a X se pueden **predecir** los valores que tomará Y. Dicho de otra manera, se puede conocer el comportamiento de Y a través de un modelo que relaciona la variación en Y con la variación de X.

El análisis de regresión tiene por objetivo **identificar** un modelo funcional que describa cómo varía la esperanza de la variable dependiente, E(Y), frente a cambios en X. Al igual que en el análisis de varianza el modelo para Y también presenta constantes desconocidas que se llaman parámetros, por lo que otro objetivo del análisis es la **estimación** de los parámetros a partir de una muestra aleatoria de observaciones en Y y en X. El análisis de regresión se ocupa también de la **validación** del modelo propuesto y de las **pruebas de hipótesis** sobre los parámetros del modelo; por último, la modelación por regresión también tiene como objetivo la **predicción**, es decir el uso del modelo para dar el valor esperado de Y cuando X toma un valor particular.

La complejidad matemática del modelo de regresión y la adecuación de éste dependerá de cuánto se conoce acerca del proceso o fenómeno que se está estudiando. En la práctica es posible adoptar modelos de regresión que se pueden agrupar o clasificar en **lineales** y **no lineales**. Los primeros hacen referencia a aquellos modelos en que la función adopta la forma de una **suma de términos**, cada uno conformado por el producto de un parámetro y una variable independiente. Los modelos no lineales son aquellos donde los parámetros no se encuentran multiplicando a las variables independientes como en el modelo lineal de tal forma que no pueden ser estimados resolviendo un sistema de ecuaciones lineales. Por ejemplo, los parámetros pueden encontrarse como **exponentes** de las variables independientes. La estimación de los parámetros en modelos no lineales se realiza usando herramientas diferentes a las presentadas en este capítulo. Aquí se abordan solamente los modelos lineales, no sólo por ser más simples, sino porque permiten dar respuesta a un gran número de problemas en las Ciencias Agropecuarias. Además, algunos de los modelos no lineales pueden, mediante adecuadas transformaciones, ser expresados de la forma lineal (en estos casos los modelos se dicen intrínsecamente lineales).

El modelo de regresión lineal más sencillo es el que se presenta en la siguiente definición:

Definición 9.1: Modelo de regresión lineal simple

Se llama **modelo de regresión lineal simple** a:

$$Y_{ij} = \alpha + \beta X_i + \varepsilon_{ij}$$

donde:

Y_{ij}= observación de la variable dependiente bajo el i-ésimo nivel de X, $i = 1,...,K$ en la j-ésima unidad experimental, $j = 1,...,m$

X_i= i-ésimo valor de la variable independiente, $i = 1,...,K$

α = parámetro que representa la ordenada al origen de la recta (indica valor esperado de Y cuando X=0)

β = parámetro que representa la pendiente de la recta (tasa de cambio en Y frente al cambio unitario en X).

ε_{ij} = variación aleatoria (o no explicada por el modelo) asociada a la j-ésima observación de Y bajo el nivel X_i.

Los ε_{ij} se suponen normales e independientemente distribuidos con esperanza 0 y varianza constante σ^2 para todo X en un intervalo donde el modelo se supone verdadero. Esto es $\varepsilon_{ij} \sim N I D\ (0, \sigma^2)$.

El modelo anterior incluye solamente una variable independiente y establece que la esperanza de la variable dependiente cambia con tasa constante, según crece o decrece el valor de la variable independiente.

¿Qué se puede decir de la esperanza de Y?, es decir ¿cuál es el valor esperado de Y para un determinado valor de X? Tomando esperanza de Y_{ij} se tiene, por propiedades de la función esperanza que:

$$E(Y_{ij} \mid X = x_i) = \mu_{y|x} = \alpha + \beta x_i$$

donde: $\mu_{y|X=x}$ representa la $E(Y_{ij})$ dado un valor de X_i, es decir la esperanza de la distribución de Y correspondiente a un valor particular de X.

α y β representan los parámetros del modelo y debe observarse que, dados α y β la esperanza de Y depende solo de X.

Cuando el investigador trata con problemas de dos variables que están ligadas por una relación funcional lineal, difícilmente los pares de observaciones (X,Y) coincidan exactamente con una recta. La presencia de errores aleatorios en las observaciones hace imposible que en la práctica se encuentre una relación funcional perfecta entre las variables. Por ello, los modelos determinísticos son de limitado valor en la descripción de fenómenos biológicos.

El modelo estadístico, a diferencia del modelo determinístico, considera una componente aleatoria con la cual se tiene en cuenta la variación de los valores de Y observados para un mismo nivel de X. Es importante notar que de la Definición 9.1 se desprende que la E(Y) se relaciona funcionalmente con X a través de una recta, luego, aún cuando las observaciones experimentales no puedan alinearse sobre la recta, si la relación funcional entre las variables existe, se espera que ésta se visualice con mayor claridad sobre los promedios.

Ejemplo 9.1

Suponga que se quiere estudiar la distribución de los pesos de una población de plantas en relación a sus alturas. Para cualquier altura elegida, por ejemplo X=50 cm, existe una distribución de pesos, es decir, la distribución de los pesos de todas las plantas de la especie que poseen esa altura. Esa distribución, llamada **distribución condicional** de Y dada X (Y|X=50), tiene como esperanza a $\mu_{Y|X=x}$ = peso medio de todas las plantas que tienen altura 50 cm y una varianza $\sigma^2_{Y|X=x}$ = varianza de los pesos de todas las plantas que tienen dicha altura.

Así, se dice que la "regresión del peso sobre la altura" representa la esperanza de la distribución de los pesos según la altura. Obsérvese la siguiente figura.

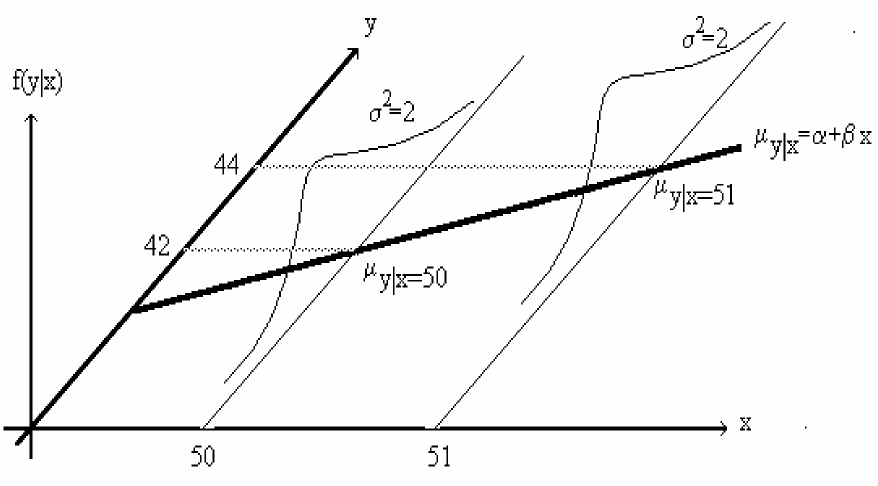


Figura 9.2: Esperanza de Y condicionada a X en relación a X.

¿Cómo se interpretan los parámetros del modelo de regresión lineal simple?

La ecuación de cualquier recta puede ser escrita como $Y = \alpha + \beta x$ donde α es la ordenada al origen e indica el valor de **y** para **x** = 0 y β es la pendiente e indica cuánto cambia **y** por cada incremento unitario en **x**. Cuando β es un número positivo significa que hay un *crecimiento* de β unidades en **y** por cada *incremento de una unidad* en **x**; si β es un número *negativo*, **y** *disminuirá* β unidades con cada incremento unitario de **x**. Luego, la pendiente y la ordenada al origen determinan la posición de la recta. En la Figura 9.3 se observa una recta con β>0.

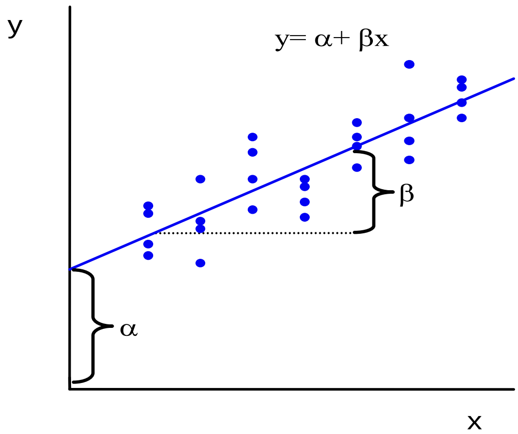


Figura 9.3: Representación gráfica de la ecuación de la recta $Y = \alpha + \beta x$ que puede describir razonablemente bien la nube de puntos presentada.

Volviendo al modelo estadístico de regresión lineal simple:

a) el parámetro α, u ordenada al origen de la recta de regresión de Y sobre X, es la esperanza de Y para X = 0; y

b) el parámetro β, o pendiente de la regresión de Y sobre X, es la diferencia entre $\mu_{Y|X=x_1}$ y $\mu_{Y|X=x_2}$ cuando x_2-x_1 = 1.

Estimación de la recta de regresión. Método de los mínimos cuadrados

Ejemplo 9.2

En un ensayo sobre trigo que se lleva a cabo en la zona de Marcos Juárez se desea cuantificar la relación que hay entre la disponibilidad de Nitrógeno en el suelo y la cantidad de Nitrógeno en la planta (que se supone lineal). Se obtuvieron datos para 12 parcelas, en las que se registró el contenido de nitrógeno en el suelo (X) y los valores promedios de nitrógeno por planta (Y). Los resultados se presentan en la Tabla 9.1.

Tabla 9.1: Cada fila representa los valores observados sobre una unidad experimental, conformada por una parcela de 50 cm. x 50 cm., en la que se midió el Nitrógeno en el suelo y por planta calculado como promedio sobre todas las plantas de la parcela

X: Nitrógeno en Suelo (ppm)	Y: Nitrógeno en planta (ppm)
0.42	0.13
0.45	0.15
0.50	0.16
0.55	0.17
0.68	0.18
0.69	0.18
0.70	0.19
0.73	0.20
0.80	0.20
0.90	0.21
0.92	0.22
0.94	0.23

El diagrama de dispersión para los datos de esta experiencia se presenta en la siguiente figura.

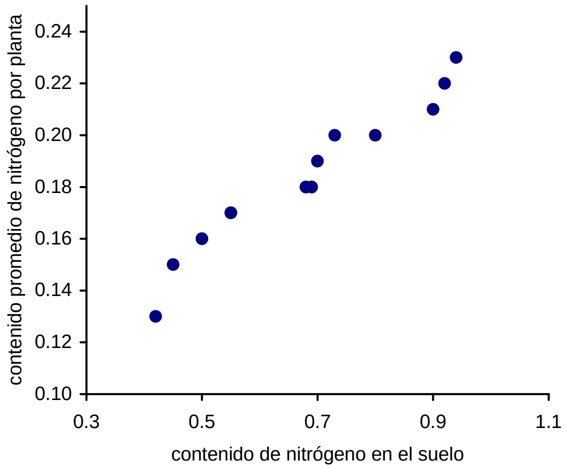


Figura 9.4: Diagrama de dispersión de los datos del Ejemplo 9.2.

El diagrama indica que hay una relación positiva entre la cantidad de nitrógeno en la planta y la cantidad de nitrógeno disponible en el suelo. En este ejemplo se puede postular una relación lineal.

La ecuación de la recta de regresión es:

$$\mu_{Y|X=x} = \alpha + \beta x$$

A partir de los datos experimentales se estiman los coeficientes α y β de la recta de regresión.

Definición 9.2: Coeficientes de regresión muestral

Se llaman **coeficientes de regresión muestral** a las estimaciones de α y β, las que se denotan como a y b respectivamente.

Si no hubiese errores aleatorios en los Y_i y el modelo lineal fuera correcto, cualquier par de puntos (X_i,Y_i) podría usarse para encontrar los valores de α y β y todas las estimaciones serían idénticas independientemente del par utilizado. Pero la presencia de los errores aleatorios descalifica este procedimiento y muestra la necesidad de disponer de un método que combine toda la información disponible en la muestra para dar una solución razonable al problema de estimación. Uno de estos métodos es el conocido como *Método de Mínimos Cuadrados*.

El método de Mínimos Cuadrados define la recta de "mejor ajuste" como aquella que hace que la suma de los cuadrados de las distancias de los valores observados respecto a la recta, medidas sobre el eje de las ordenadas, sea lo más pequeña posible. Esto es:

$$\min_{a,b} \sum_{i=1}^{n} \left(y_i - \hat{y} \right)^2 = \sum_{i=1}^{n} \left(y_i - a - b.x_i \right)^2 = \sum_{i=1}^{n} e_i^2$$

donde:

$\hat{y} = a + bx$, es el valor *predicho* por el modelo lineal y e_i es el *residuo* definido como $e_i = \left(y_i - \hat{y} \right)$.

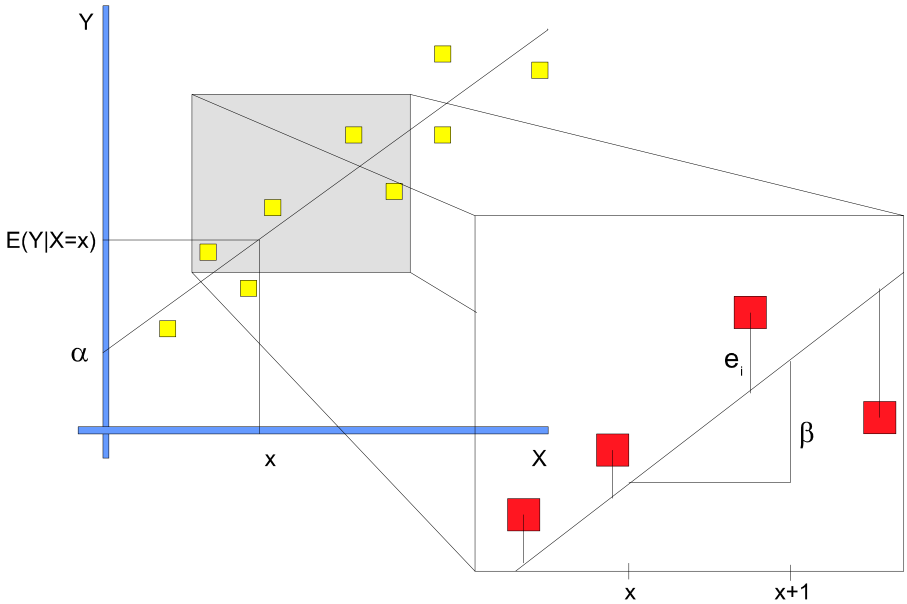


Figura 9.5: Representación de los residuos, E(Y|X=x), recta de regresión e interpretación geométrica de la ordenada al origen (α) y de la pendiente (β) de la recta

El método de estimación por mínimos cuadrados produce las siguientes expresiones para los estimadores **b** y **a** de β y α respectivamente:

$$b = \frac{\sum_{i=1}^{n} X_i Y_i \ - \ \frac{\sum_{i=1}^{n} X_i . \sum_{i=1}^{n} Y_i}{n}}{\sum_{i=1}^{n} X_i^2 \ - \ \frac{(\sum_{i=1}^{n} X_i)^2}{n}}, \qquad\qquad a = \overline{y} - \beta \overline{x}$$

En el ejemplo:

$$b \ = \ \frac{1.5888 \ - \ \frac{8.28 \ . \ 2.22}{12}}{6.0728 \ - \ \frac{8.28^2}{12}} = \ 0.159$$

$$a = 0.185 - 0.159.\ 0.69 = 0.076$$

por tanto la regresión estimada de Y sobre X puede expresarse como:

$$\mu_{\ Y \mid X = x_i} = 0.076 + 0.159\ x_i$$

y su gráfica se presenta en la Figura 9.6.

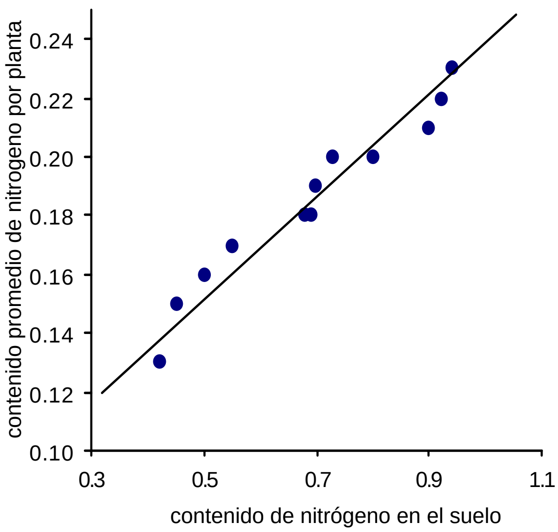


Figura 9.6: Representación gráfica conjunta del diagrama de dispersión del Ejemplo 9.2 y la recta de regresión estimadas Y = 0.076 + 0.159 X.

Estimaciones y predicciones

La ecuación de regresión puede ser usada para obtener estimaciones de la esperanza de Y o predicciones de Y para valores elegidos de X. Debe tenerse en cuenta, sin embargo, que los valores de X propuestos deben pertenecer al dominio de las X utilizado para la estimación de la recta. No es conveniente usar la ecuación de la recta para extrapolar, es decir para estimar la esperanza de Y para valores de X fuera del rango estudiado ya que no se conoce nada sobre el comportamiento de la relación de X e Y fuera del dominio en la que se estudió esta relación. Por supuesto, aún dentro del dominio estudiado de X, la validez de las estimaciones depende de la *bondad de ajuste del modelo*, es decir su grado de aproximación respecto de la verdadera relación funcional entre las variables.

Cada valor calculado a partir de la recta de regresión, es la estimación de la esperanza de la distribución de Y condicionada a un valor de X $\left(\hat{\mu}_{Y|X=x}\right)$, o una predicción del valor de Y para una observación futura de X ($\hat{y}$).

En el ejemplo, las predicciones de Y para x = 0.93 y x = 0.46 son, respectivamente:

$$\hat{y} = 0.076 + 0.159\,(0.93) = 0.22$$

$$\hat{y} = 0.076 + 0.159\,(0.46) = 0.15$$

Intervalo de confianza para la esperanza condicional de Y

Utilizando las propiedades de la varianza de la suma de variables aleatorias, aplicada a la expresión de la esperanza condicional de Y dado X se tiene:

$$Var\left(E\left(Y \mid X = x\right)\right) = \sigma^2\left[\frac{1}{n} + \frac{\left(x-\overline{x}\right)^2}{\sum x_i^2 - \left(\sum x_i\right)^2/n}\right]$$

De la expresión anterior pueden deducirse tres propiedades:

a) La varianza de la esperanza de Y no es igual para todo valor X_i, de hecho es mínima cuando X_i coincide con la media muestral de X.

b) La varianza de la esperanza de Y es más pequeña cuanto mayor es la suma de cuadrados de X $\left(\sum x_i^2 - \left(\sum x_i\right)^2/n\right)$, lo que implica que cuanto más disímiles sean los valores de X a los cuales se observan los valores de Y, tanto mejor serán las estimaciones de las esperanzas condicionales de Y.

c) Para *n* que tiende a infinito la varianza de la esperanza condicional de Y tiende a cero.

Además, bajo los supuestos clásicos del análisis de regresión, el intervalo de confianza al 95%, de μ_Y para $X=x_0$ está dado por:

$$\hat{y}_0 \pm 1.96\sqrt{\sigma^2\left[\frac{1}{n}+\frac{(x_0-\bar{x})^2}{\sum x_i^2-\left(\sum x_i\right)^2/n}\right]}$$

Si σ^2 no se conoce y se estima, entonces, el intervalo anterior se modifica reemplazando el valor 1.96 por el cuantil correspondiente de una T con n-2 grados de libertad y sustituyendo σ^2 por su estimador.

Cuando los intervalos de confianza se grafican para todos los valores de x en un recorrido dado se obtienen *bandas de confianza*. La Figura 9.7, muestra las *bandas de confianza* al 95% para una regresión lineal simple en la que se evaluó el contenido de nitrógeno en plantas de trigo en función del contenido de nitrógeno del suelo.

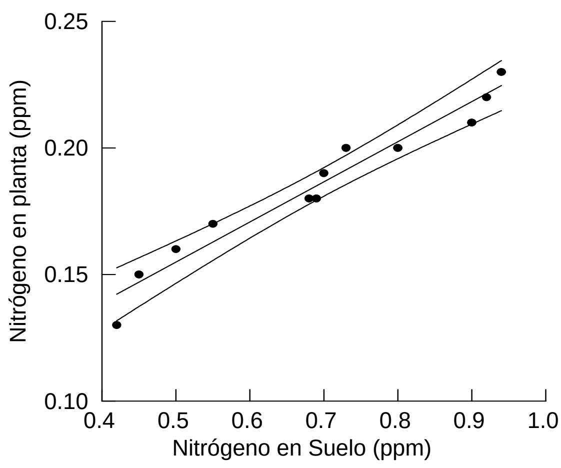


Figura 9.7: Recta de mínimos cuadrados y bandas de confianza al 95% para la esperanza condicional de Y dado X=x.

Intervalo de predicción de Y dado X

Al igual que en el punto anterior, aplicando el operador varianza al predictor de Y dado X=x se tiene la siguiente expresión.

$$Var\left(Ypred \mid X = x\right) = \sigma^2 \left[1 + \frac{1}{n} + \frac{\left(x - \bar{x}\right)^2}{\sum x_i^2 - \left(\sum x_i\right)^2 / n} \right]$$

Idénticas observaciones a las realizadas para la varianza de la esperanza condicional de Y, se pueden hacer para la expresión anterior, pero debe agregarse que en este caso la varianza es σ^2 unidades mayor y que para *n* que tiende a infinito la varianza del predictor tiende a σ^2.

Cuando se grafican todos los intervalos de predicción para una región dada de x, se obtienen las *bandas de predicción*, que son similares a las de confianza, excepto que son más amplias.

El intervalo de predicción al 95% de Y dado $X=x_0$ tiene la siguiente expresión:

$$\hat{y}_0 \pm 1.96 \sqrt{\sigma^2 \left[1 + \frac{1}{n} + \frac{\left(x_0 - \bar{x}\right)^2}{\sum x_i^2 - \left(\sum x_i\right)^2 / n} \right]}$$

En el caso en que se estime σ^2, el intervalo se obtiene reemplazando 1.96 por el cuantil correspondiente de una T con n-2 grados de libertad y sustituyendo σ^2 por su estimador.

La diferencia entre intervalo de confianza y predicción esta dada en que el primero delimita una región que con probabilidad 1-α contiene a la verdadera esperanza de Y dado X, mientras que el segundo delimita un región cuya probabilidad de ocurrencia para muestras aleatorias de Y dado X es 1-α.

Intervalo de confianza para la ordenada al origen

Para dar un intervalo de confianza para la ordenada al origen del modelo de regresión lineal simple se necesita conocer la varianza del estimador "a" de α. La siguiente expresión de la varianza de "a" se obtiene aplicando las reglas del operador varianza al estimador de α:

$$Var(a) = \left(\frac{1}{n} + \frac{\left(\bar{x}\right)^2}{\sum x_i^2 - \left(\sum x_i\right)^2 / n} \right) \sigma^2,$$

donde σ^2 es la varianza del error. Dado que bajo los supuestos usuales de regresión "a" se distribuye como una normal con esperanza α y varianza según la expresión anterior, el intervalo de confianza al 95% para α esta dado por:

$$a \pm 1.96\sqrt{\left(\frac{1}{n}+\frac{(\bar{x})^2}{\sum x_i^2-\left(\sum x_i\right)^2/n}\right)\sigma^2}$$

Si no se conoce σ^2 y se estima, como se verá más adelante, entonces el intervalo se obtiene utilizando el cuantil correspondiente de una T con n-2 grados de libertad en reemplazo de 1.96 y sustituyendo σ^2 por su estimador.

Intervalo de confianza para la pendiente

Al igual que para la ordenada al origen, la obtención de un intervalo de confianza para β se basa en la distribución de su estimador "b" y la varianza del mismo. Bajo los supuestos que se tienen para el análisis de regresión, "b" se distribuye normal con esperanza β y varianza dada por la siguiente expresión:

$$Var(b)=\left(\frac{\sigma^2}{\sum x_i^2-\left(\sum x_i\right)^2/n}\right)$$

donde σ^2 es la varianza del error. Luego, el intervalo de confianza al 95% para β esta dado por:

$$b \pm 1.96\sqrt{\left(\frac{\sigma^2}{\sum x_i^2-\left(\sum x_i\right)^2/n}\right)}$$

Si no se conoce σ^2 y se estima, entonces el intervalo se obtiene sustituyendo 1.96 por el cuantil correspondiente de una T con n-2 grados de libertad y σ^2 por su estimador.

Pruebas de hipótesis en regresión

En los puntos anteriores se ha estudiado como estimar los parámetros de un modelo de regresión lineal simple: estos son la ordenada al origen (α) y la pendiente (β). En esta sección se aborda la problemática de la prueba de hipótesis sobre estos parámetros. La aproximación más simple para probar $\alpha = \alpha_0$ y/o $\beta = \beta_0$ es mediante un test T. Los estadísticos de las pruebas T, que se presentan a continuación, son simples y bajo los supuestos, que se discutirán más adelante, se distribuyen como una T con n-2 grados de libertad.

Para pruebas de hipótesis sobre α	Para pruebas de hipótesis sobre β
$T = \dfrac{a-\alpha_0}{\sqrt{\left(\dfrac{1}{n}+\dfrac{(\bar{x})^2}{\sum x_i^2-\left(\sum x_i\right)^2/n}\right)\hat{\sigma}^2}}$	$T = \dfrac{b-\beta_0}{\sqrt{\left(\dfrac{\hat{\sigma}^2}{\sum x_i^2-\left(\sum x_i\right)^2/n}\right)}}$

En las expresiones dadas aparece la estimación de la varianza del error $\left(\hat{\sigma}^2\right)$. No se ha mostrado, hasta ahora, una expresión para este estimador, sin embargo, ésta no es desconocida ya que se presentó en el contexto del análisis de la varianza. La técnica de estimación nos conduce a la partición de la Suma de Cuadrados Total (SCT) de Y en una Suma de Cuadrados Explicada por α (SCα), una Suma de Cuadrados Explicada por β (SCβ) y una Suma de Cuadrados Residual (SCR). Así, se tiene:

$$\text{SCT} = \text{SC}\alpha + \text{SC}\beta + \text{SCR}$$

Las sumas de cuadrados dadas tienen grados de libertad asociados. Las SCα y SCβ tienen ambas 1 grado de libertad cada una, la SCT tiene "n" y SCR "n-2". Luego, $\hat{\sigma}^2 = \text{SCR/(n-2)}$. La descomposición de la suma de cuadrados permite estimar $\hat{\sigma}^2$ y construir la siguiente tabla de ANOVA para el modelo de regresión:

Fuentes de Variación	Suma de Cuadrados	Grados de libertad	Cuadrados Medios	F
α	SCα	1	CMα	CMα /CMR
β	SCβ	1	CMβ	CMβ/CMR
CMR	SCR	n-2	CMR	
Total	SCT	n		

Las pruebas F de las dos primeras filas de la tabla sirven para probar las hipótesis: H_0:α=0 vs H_1: α≠0 y H_0:β=0 vs H_1:β≠0 respectivamente. Es usual que la prueba H_0:α=0 sea irrelevante o carente de sentido en el contexto del problema y la presencia de α en el modelo cumple sólo con el propósito de no poner restricciones al ajuste lineal. Por lo tanto, virtualmente todo el *software* estadístico omite la prueba H_0: α= 0 y en el caso de proveer el cálculo de la SCT, lo que muestran es una SCT corregida que es igual SCT-SCα con "n-1" grados de libertad. Debido a que la corrección de la SCT es la práctica usual, excepto que se indique lo contrario, siempre se hace referencia a ella. De esta forma SCT $_{(corregida)}$ = SCβ + SCR y la tabla de ANAVA es la siguiente:

Tabla 9.2: Cuadro de Análisis de la Varianza para la hipótesis usual del modelo de regresión simple. H_0: $\beta = 0$, siendo β el coeficiente

Fuentes. de variación	Suma de Cuadrados	Grados de libertad	Cuadrados Medios	F observada
Debida a β (explicada)	$\dfrac{\left(\sum_{i=1}^{n} X_i Y_i - \dfrac{\sum_{i=1}^{n} X_i \sum_{i=1}^{n} Y_i}{n}\right)^2}{\sum_{i=1}^{n} X_i^2 - \dfrac{\left(\sum_{i=1}^{n} X_i\right)^2}{n}}$	1	$\dfrac{SC\beta}{1}$	$\dfrac{CM\beta}{CMR}$
Residual (no explicada)	SC Total-SCβ	n-2	$\dfrac{SCR}{n-2}$	
Total (corregida)	$\sum_{i=1}^{n} Y_i^2 - \dfrac{(\sum_{i=1}^{n} Y_i)^2}{n}$	n-1		

Observación: Como podrá observarse, la suma de cuadrados total (corregida) es idéntica a la que se encontró en el análisis de la varianza mientras que el Cuadrado Medio Residual es el estimador de la varianza del error (σ^2) al igual que en el análisis de la varianza lo era la suma de cuadrados del error. La SCβ es también conocida como *Suma de Cuadrados de Regresión*.

Ejemplo 9.2: (continuación) volviendo a la relación entre el contenido de Nitrógeno en planta y en suelo presentada anteriormente y después de obtener las estimaciones de α y β, se puede proceder con la prueba de hipótesis para establecer el rechazo o no de la hipótesis $\beta = 0$.

Los cálculos para el ejemplo son:

$$\text{SCTotal} = 0.4202 - \frac{2.22^2}{12} = 0.0095$$

$$SC\beta = \frac{\left(1.5888 - \frac{8.28 \cdot 2.22}{12}\right)^2}{6.0728 - \frac{8.28^2}{12}} = \frac{0.057^2}{0.3596} = 0.0090$$

$$\text{SCR} = \text{SCT} - \text{SC}\beta = 0.0095 - 0.0090 = 0.0005$$

Tabla 9.3: Tabla del Análisis de Regresión del Ejemplo 9.2

Fuentes de variación	Suma de Cuadrados	GL	Cuadrados Medios	F Observada
Debida a β (explicada)	0.0090	1	0.0090	180
Residual (no explicada)	0.0005	10	0.00005	
Total (corregido)	0.0095	11		

Como la F observada es mayor que el cuantil (1-α) de una $F_{1,10}$ se rechaza H_0 y se concluye que un modelo lineal para la relación entre nitrógeno en la planta y nitrógeno en el suelo explica una parte de la variación del contenido de Nitrógeno en la planta que resulta estadísticamente significativa.

Si la hipótesis nula se acepta, no puede asegurarse que la pendiente de la recta de regresión estimada sea diferente de cero. Luego, si la recta tiene pendiente nula, los valores de Y son indiferentes a los valores de X y por lo tanto la relación lineal propuesta no explica las variaciones de Y en función de X.

Los supuestos del análisis de regresión

Tanto los métodos de estimación de los parámetros del modelo de regresión, así como los intervalos de confianza hallados y las pruebas de hipótesis estudiadas son válidas si se cumplen las siguientes propiedades estadísticas para los errores del modelo.

La *esperanza de la distribución de los errores es 0*: E (ε_i) = 0 $\forall$i

La *varianza de la distribución de los errores es constante:* V (ε_i) = σ^2 $\forall$i

Los ε_i *son variables aleatorias normales e independientes*.

Estas tres propiedades se resumen indicando que $\varepsilon_i \sim$ NIID (0, σ^2) y que se lee: los errores son variables aleatorias normales independientes e idénticamente distribuidas con esperanza 0 y varianza σ^2.

Además, de los supuestos sobre los errores, también se supone válido el modelo lineal para la esperanza condicional de Y. Es decir, se supone cierto que $E(Y|X= x) = \alpha+ \beta x$.

El análisis de regresión está estrechamente ligado al análisis de la varianza y los supuestos son los mismos para ambas técnicas. En ambos casos los supuestos soportan las propiedades estadísticas que hacen válida la inferencia. Si los supuestos no se cumplen, el método de estimación por mínimos cuadrados no es necesariamente el más eficiente, los intervalos de confianza hallados, el nivel de significación y potencia nominales de las pruebas estadísticas de hipótesis no coinciden con sus verdaderos valores. Es por esta razón útil preguntarse sobre la razonabilidad de los supuestos en cada problema real y en caso necesario validarlos a través de pruebas gráficas o formales. Si alguno de los supuestos no se cumple usualmente se transforman los datos originales llevándolos a una escala en la que los supuestos se cumplen. Otra alternativa es usar métodos estadísticos que no exigen el cumplimiento de estos supuestos.

Valor predictivo del modelo de regresión

Se ha indicado que la variación total en Y puede ser vista como la variación explicada por la regresión más la variación no explicada o residual. Si la variación no explicada es substancialmente mayor que la variación explicada, se tendrá un indicio de que modelo no es bueno para fines predictivos, es decir, el modelo está explicando poco de la variación en Y. No se debe, sin embargo, confundir la medida de cuanto explica un modelo con su pertinencia, ya que se recordará una vez más, que el modelo es para las esperanzas de Y. Una medida muestral de la capacidad predictiva del modelo es el coeficiente de determinación, denotado por R^2.

Definición 9.3: Coeficiente de determinación muestral

Llamaremos **coeficiente de determinación** muestral a:

$$R^2 = \frac{\text{Suma de Cuadrados de Regresión}}{\text{Suma de Cuadrados Total}}$$

Este coeficiente se interpreta como la proporción de la *variabilidad* total en Y explicable por la variación de la variable independiente o como también es usual decir: la proporción de la variabilidad total *explicada por el modelo*. Por ser una proporción, el coeficiente de determinación varía entre 0 y 1. Cuanto más próximo esté a 1, mayor valor predictivo tendrá el modelo en el sentido que los valores observables estarán muy próximos a la esperanza estimada por la regresión.

Siguiendo con el ejemplo de la relación entre Nitrógeno en planta y Nitrógeno en suelo, el coeficiente de determinación obtenido es $R^2 = 0.951$, es decir el 95% de la suma de cuadrados totales de la variable dependiente (Nitrógeno en planta) es "explicada", a través de una relación lineal, por la variación observada en la variable independiente.

Es frecuente ver al coeficiente de determinación usado como una medida de la adecuación del modelo, entendiendo por adecuación que la relación funcional y los supuestos sobre los errores son correctos. Esta interpretación es **absolutamente incorrecta** y se pueden dar ejemplos en los que R^2 es muy alto y el modelo completamente inapropiado. Luego, R^2 es válido como medida de ajuste o de valor predictivo si el modelo es correcto tanto en su parte determinística como en su parte aleatoria. La evaluación de la adecuación del modelo es un tema amplio que excede el objetivo de este libro pero es una de las áreas a las que se ha prestado mucha atención en los últimos años y existe una amplia bibliografía sobre el tema (Rawlings, 1988, Myers,1990; Draper y Smith, 1998)

Análisis de Correlación Lineal

En el análisis de regresión, la variable X es usualmente fija, mientras que la variable dependiente Y es aleatoria. Si X e Y son **ambas** variables aleatorias observables sobre una misma unidad o elemento de la población, podría ser de interés medir el grado en que estas variables covarian ya sea positiva o negativamente. Por ejemplo, si un fitomejorador sabe cómo controlar la altura del tallo de maíz y se puede establecer que existe un alto grado de asociación entre la altura del tallo y el rendimiento de la cosecha se podrá, probablemente, también controlar el rinde.

La simple observación de que dos variables parecen estar relacionadas, no revela gran cosa. Dos importantes preguntas se pueden formular al respecto:

a) ¿Qué tan estrechamente relacionadas se encuentran las variables? o ¿cuál es el grado de asociación que existe entre ambas?

b) ¿Es real la asociación observada o podría haber ocurrido solo por azar?

Para responder a la primer pregunta se necesita una medida del grado de asociación entre las dos variables. Esta medida es el **coeficiente de correlación,** que se denota con la letra griega ρ (rho). Para la segunda, se precisa una prueba estadística de hipótesis para ρ.

El análisis de correlación clásico supone que los pares (X_i, Y_i) son pares de variables aleatorias idénticamente distribuidos con **distribución normal bidimensional, o normal bivariada**. Geométricamente, la función de densidad de esta distribución es una superficie de forma acampanada. La distribución normal bivariada es aquella en la que la distribución condicional de Y para cualquier X, es normal, y la distribución condicional de X para cualquier Y, es también normal. Esta distribución incluye a ρ como uno de sus parámetros. Las siguientes figuras muestran una normal bivariada con $\rho = 0$ y una normal bivariada con $\rho = 0.8$.

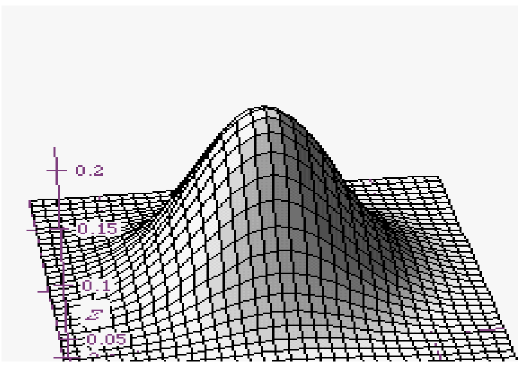


Figura 9.8:Densidad normal bivariada: ρ=0.

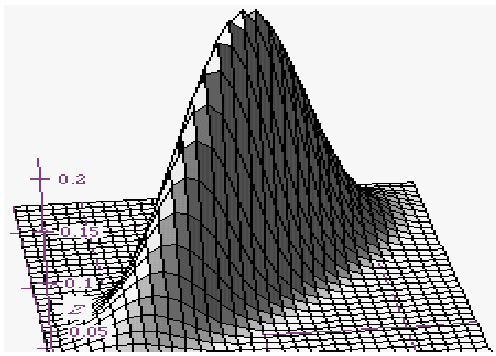


Figura 9.9:Densidad normal bivariada: ρ=0.8.

Observación: Aunque en el análisis de correlación no se explicita la forma de la asociación entre variables cuya intensidad y sentido se quiere medir, el coeficiente de correlación clásico o de Pearson cuantifica el **grado de asociación lineal** entre ellas. Por lo tanto si dos variables siguen una estrecha asociación **no lineal**, el coeficiente de correlación no la cuantificará correctamente.

Definición 9.4: Coeficiente de correlación lineal.

El **coeficiente de correlación lineal** entre las variables aleatorias X e Y se define como :

$$\rho = \frac{\mathrm{cov}(X,Y)}{\sqrt{Var(X)Var(Y)}}$$

donde Var(X) y Var(Y) denotan las varianzas de X e Y respectivamente y Cov(X,Y) denota la covarianza entre X e Y que se define como Cov(X,Y)= E (XY) - E(X) E(Y).

Es importante observar que de la definición surge que el coeficiente de correlación es independiente de las unidades de medida de las variables. También debe notarse que el coeficiente de correlación lineal *vive* en el intervalo [-1,1].

Este coeficiente es un indicador de la densidad alrededor de la recta de regresión para la distribución condicional de Y dado X y viceversa.

Cuando X e Y están no correlacionadas, ρ es igual a cero. En este caso el conocimiento de una de las variables no ayuda a describir el comportamiento de la otra. Por otra parte, cuando X e Y están altamente correlacionadas en forma lineal, ρ está muy próximo a 1 ó -1.

Por definición de la normal bivariada, ρ es un parámetro que la caracteriza, y como todo otro parámetro, se estima a partir de observaciones muestrales.

Definición 9.5: Coeficiente de correlación lineal muestral de Pearson

Si $(X_1, Y_1),..., (X_n, Y_n)$ es una muestra aleatoria bivariada de tamaño n, el **coeficiente de correlación lineal muestral** (estimador de ρ), se denota con **r** y se define por:

$$r = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^{n}(X_i - \bar{X})^2 \sum_{i=1}^{n}(Y_i - \bar{Y})^2}} .$$

La fórmula de cálculo es:

$$r = \frac{\sum_{i=1}^{n} X_i Y_i - \frac{\sum_{i=1}^{n} X_i \sum_{i=1}^{n} Y_i}{n}}{\sqrt{\left(\sum_{i=1}^{n} X_i^2 - \frac{\left(\sum_{i=1}^{n} X_i\right)^2}{n}\right)\left(\sum_{i=1}^{n} Y_i^2 - \frac{\left(\sum_{i=1}^{n} Y_i\right)^2}{n}\right)}}$$

Este estimador provee una medida muestral de la correlación entre X e Y, y posee la propiedad de ser un estimador insesgado de ρ cuando ρ = 0.

Cuando ρ está en la proximidad de 1 o -1 los pares (x,y) se alinean sobre una recta con pendiente positiva o negativa según el signo del coeficiente. Cuando ρ = 0, los pares (X,Y) están dispersos alrededor del punto $\left(\overline{X},\overline{Y}\right)$ sin ninguna dirección predominante.

nota: ρ ≠ 0 implica solamente que hay asociación entre X e Y pero no implica relaciones de causalidad. Bajo el supuesto de distribución normal bivariada ρ = 0 implica que X e Y son estadísticamente independientes.

Prueba de hipótesis sobre ρ

Si se satisfacen las suposiciones de normalidad bivariada y se tiene una muestra aleatoria de n pares de valores (X,Y), es posible utilizar el coeficiente de correlación muestral "r", para probar la independencia entre X e Y probando la hipótesis H_0:ρ = 0.

Para probar la hipótesis H_0:ρ = 0 vs. H_1: ρ ≠ 0, el estadístico utilizado es:

$$T = \frac{r}{\sqrt{\frac{1-r^2}{n-2}}}$$

que se distribuye como una distribución T de Student con n-2 grados de libertad, donde n es el número de pares (X,Y). Luego se procede como en cualquier prueba de hipótesis para la aceptación o rechazo de H_o.

Ejemplo 9.3

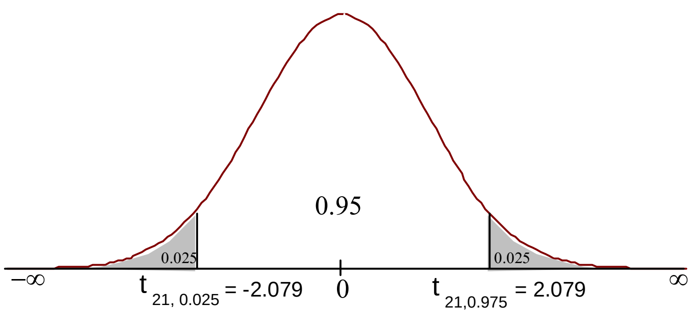
Los datos de la Tabla 9.4 se refieren al contenido de proteína bruta (PB) y caseína (CA) en leche en una muestra de 23 tambos de la cuenca lechera del centro del país.

Tabla 9.4: Contenido de proteína bruta (PB) y caseína (CA) en leche de 23 tambos de la cuenca lechera de la región central Argentina.

PB	*CA*	*PB*	*CA*
2.74	1.87	2.95	2.04
3.19	2.26	3.08	2.16
2.96	2.07	3.14	2.16
2.91	2.09	3.22	2.22
3.23	2.28	3.14	2.22
3.04	2.04	3.15	2.24
3.08	2.18	3.2	2.22
3.23	2.3	2.95	2.07
3.11	2.17	3.19	2.25
3.11	2.15	3.12	2.23
3.1	2.16	2.99	2.16
3.25	2.33		

El coeficiente de correlación lineal muestral entre PB y CA es: r = 0.9327. ¿Es esta alta correlación estadísticamente significativa? Para contestar a esta pregunta se debe realizar una prueba de hipótesis:

Las hipótesis en este caso son: H_0: $\rho = 0$ vs H_1: $\rho \neq 0$. Fijando $\alpha = 0.05$ y utilizando el estadístico $T = \dfrac{r}{\sqrt{\dfrac{1-r^2}{n-2}}}$, que se distribuye bajo H_0 como una T de Student con n-2 grados de libertad, se determina la región de aceptación como el intervalo delimitado por los cuantiles 0.025 y 0.975 de una $t_{(n-2)}$ como se muestra en la siguiente figura.



Calculando el estadístico se tiene $T = \dfrac{0.9327}{\sqrt{\dfrac{1-0.9327^2}{23-2}}}$ = 11.85, que está fuera de la región de aceptación y por lo tanto se rechaza H_0. Se concluye luego que, con un nivel de significación del 5%, se rechaza la hipótesis de correlación nula. En consecuencia se puede decir que hay una correlación lineal estadísticamente significativa entre los porcentajes de proteína bruta y caseína en la leche.

Ejercicios

Ejercicio 9.1

Los siguientes datos corresponden a los porcentajes de mortalidad obtenidos a dosis crecientes de un insecticida. Se desea estudiar si existe una componente lineal entre la mortalidad y la dosis, expresada como el logaritmo de las concentraciones utilizadas. El experimento consistió en someter a grupos de 1000 insectos a cada una de las dosis ensayadas. Los resultados fueron los siguientes:

Ln(dosis)	Mortalidad (%)
0	5
1	7
5	10
10	16
15	17
20	25
25	26
30	30

a) Construir un diagrama de dispersión Mortalidad vs. Ln(dosis).

b) De acuerdo al gráfico obtenido, ¿es razonable proponer un ajuste lineal?

c) Escribir el modelo lineal que, se supone, relaciona la mortalidad con la dosis.

d) Estimar los parámetros del modelo.

e) Construir el cuadro de análisis de la varianza y obtener conclusiones.

Ejercicio 9.2

Considérese nuevamente un ensayo para evaluar el efecto comparativo de dos insecticidas (A y B) sobre la mortalidad de insectos. Con los resultados que se presenta a continuación:

	Mortalidad (%)	
Ln(dosis)	Insecticida A	Insecticida B
0	5	6
1	7	5
5	10	8
10	16	8
15	19	13
20	27	17
25	28	22
30	34	23

a) Verificar si para los insecticidas "A" y "B" es *razonable* un modelo lineal de la forma $Y = \alpha + \beta x + \varepsilon$ para modelar la mortalidad en relación a la dosis.

b) Estimar los parámetros de ambos modelos.

c) Construir los cuadros de análisis de la varianza.

d) Comparar las pendientes y ordenadas al origen de ambos insecticidas.

e) Si el ensayo ha sido bien planificado, ¿qué se espera de la diferencia de las ordenadas al origen?

f) ¿Qué se recomienda teniendo en cuenta las pendientes?

Ejercicio 9.3

Para estudiar el efecto de la temperatura sobre el vigor durante la germinación, se dispusieron semillas de alfalfa en germinadores a distintas temperaturas. A los 6 días se midió la longitud de las plántulas, obteniéndose los siguientes datos:

T (^oC)	Longitud de Plantas (mm)					
10	13	18	15	19	11	17
15	20	24	15	17		
20	22	27	31	21	26	
25	24	25	28	23		

a) ¿Qué diferencia hay en los datos de este ejercicio con respecto a los anteriores?

b) Construir el diagrama de dispersión entre longitud de plántula y temperatura y verificar si existe una tendencia lineal.

c) Realizar un análisis de regresión lineal trabajando con $\alpha = 0.05$.

d) ¿Qué temperatura permite obtener mayor vigor?

Ejercicio 9.4

Si los rendimientos del ajo dependen linealmente, en un cierto rango, del porcentaje de materia orgánica (MO) del suelo con pendiente 4000kg/ha/MO(%), ¿cuál es la diferencia promedio de rendimiento entre campos que poseen una diferencia en el contenido de materia orgánica del suelo del 1.3%? (Se supone que estos campos tienen contenidos de materia orgánica en el rango de validez del modelo y que el modelo es válido en ambos campos).

Ejercicio 9.5

En un experimento para evaluar la efectividad de un insecticida sobre la sobrevida de dos especies de insectos (A y B) se obtiene que, en ambos casos, es posible ajustar un modelo lineal para la sobrevida (Y) versus la concentración (en ppm) del insecticida utilizado (X), siendo los modelos ajustados los siguientes:

$$\textit{Especie A: } Y = 80 - 15X; \qquad \textit{Especie B: } Y = 60 - 15X.$$

De acuerdo a estos resultados:

a) ¿Es el insecticida igualmente efectivo en ambas especies?

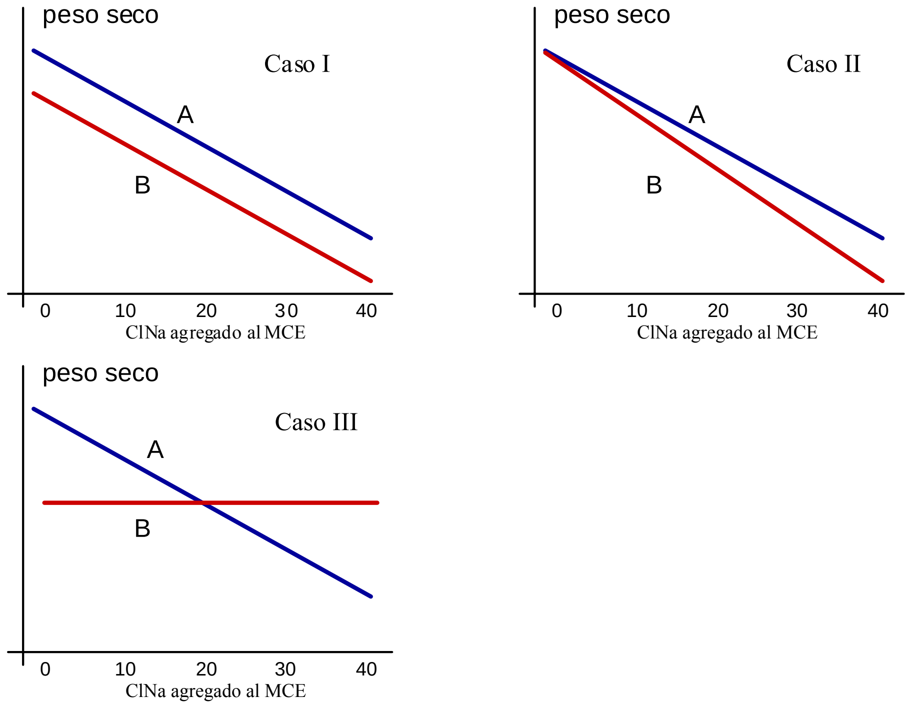
b) ¿Qué interpretación se puede hacer de cada una de estas ecuaciones?

c) ¿Cómo se modifica la sobrevida por cada incremento unitario en la concentración del insecticida agregado?

d) Si se quisiera que ambas especies tengan una sobrevida de a lo sumo 20, ¿cuántas pm. se debería agregar del insecticida?

Ejercicio 9.6

En un ensayo de resistencia a la sequía, dos especies de leguminosas (A y B) fueron comparadas. El experimento consistió en registrar el peso seco total de 10 plantas al cabo de 30 días desde la siembra. Las condiciones comparadas fueron las siguientes: medio de cultivo estándar (MCE), MCE+10 g/l de ClNa, MCE+20 g/l de ClNa, MCE+30 g/l de ClNa, MCE+40 g/l de ClNa. Los siguientes tres gráficos muestran tres resultados posibles para esta experiencia. Los gráficos representan las rectas que modelan la esperanza del peso seco en relación al agregado de ClNa en cada caso.



a) ¿Qué conclusión se obtendría, en cada una de estas situaciones acerca de la resistencia a la sequía de ambas especies, asumiendo que si la especie soporta mayor contenido de ClNa será más resistente?

b) ¿Qué significan (o que interpretación tienen) la diferencia y la similitud de las ordenadas al origen de las rectas ajustadas en los casos I, II, y III?

c) ¿Qué significan (o que interpretación tienen) la diferencia y la similitud de las pendientes de las rectas ajustadas en los casos I, II, y III?

Ejercicio 9.7

Se desea probar la efectividad de un nuevo fungicida para el control de roya en trigo. Se probaron distintas dosis en gramos de principio activo por ha (gr.p.a./ha) en 10 parcelas de 100 plantas cada una. A los 15 días de la aplicación se realizó una evaluación del daño, como el tamaño promedio de las machas en hoja bandera. Los datos son los siguientes:

Dosis(X)	100	125	200	250	275	300	325	350	375	400
Daño (Y)	50	48	39	35	30	25	20	12	10	5

a) Ajustar un modelo de regresión lineal para el daño en función de la dosis y construir las bandas de predicción y de confianza.

b) Predecir el daño (tamaño promedio de las manchas) que se hallará si se aplican 260 gr.p.a./ha

Ejercicio 9.8

En un estudio se hicieron mediciones de perímetro y peso de cabezas de ajo. Los datos que se obtuvieron fueron los siguientes:

Perímetro (cm)	12.39	12.39	12.71	9.8	12.3	10.12	11.81	11.41	9.4	11.49
Peso (grs.)	32.27	29.39	30.8	15.6	29.8	16.87	28.11	23.29	14.11	25.37

a) ¿Cómo se espera que sea la asociación entre peso y perímetro?

b) Calcular coeficiente correlación entre peso y perímetro

c) ¿Es significativo el coeficiente encontrado?

d) Elaborar conclusiones.

10

Diseño de Experimentos

Introducción

El objetivo de este capítulo es dar un panorama de los principios y técnicas del diseño de experimentos. No pretende dar respuesta a la amplia variedad de situaciones experimentales ni hacer un recuento de las técnicas disponibles para abordar estas situaciones. Por el contrario en este material se presenta una selección de tópicos que permite al lector comprender los fundamentos del diseño y abordar la lectura de textos mas avanzados y completos.

Elementos del Diseño de Experimentos

Para abordar el tratamiento de los tópicos del diseño de experimentos, se presentarán a continuación las definiciones necesarias para el desarrollo y discusiones posteriores.

Experimento

Definición 10.1: Experimento

Se define a un **experimento** como la acción de aplicar uno o más tratamientos a un conjunto de unidades experimentales para valorar sus respuestas.

Bajo el modelo experimental, las alteraciones en las respuestas se atribuyen solamente a la acción de los tratamientos excepto por variaciones aleatorias (usualmente pequeñas) debidas a errores experimentales y/o falta de homogeneidad de las unidades experimentales.

Unidad experimental

La definición de parcela o unidad experimental se dio en el Capítulo 8, por lo que solo se recuerda que se trata de una alícuota de material, una parcela de terreno, un animal o grupo de animales, etc. al cual se le aplica un tratamiento y sobre el que, posteriormente, se observan una o más respuestas para evaluar el efecto del tratamiento.

Factores y Tratamientos

Los tratamientos que reciben las unidades experimentales pueden corresponder a distintas dosis de una droga, a diferentes intensidades de luz, a cantidades variables de agua o a distintos tipos de insecticidas. En cada uno de estos casos, se dice que el experimento es **unifactorial** ya que los tratamientos consisten en aplicar distintos **niveles** de un mismo **factor**. Un ejemplo de estos experimentos es aquel en que se ensayan distintas densidades de siembra para evaluar los rendimientos agrícolas. En este ejemplo la densidad de siembra es el factor y las distintas densidades sus niveles.

Si en cambio un tratamiento consiste en la **combinación de niveles** de 2 o más factores, entonces se dice que el experimento es un experimento **factorial**. A modo de ejemplo, suponga que en un ensayo comparativo de rendimientos se siembran parcelas experimentales con tres variedades de una especie en dos fechas de siembra. En este ensayo, se tienen dos factores: *variedad* y *fecha de siembra* y un total de seis tratamientos, el factor variedad se encuentra a tres niveles y el factor fecha a dos.

Modelo para las observaciones

A los fines del tratamiento estadístico de los resultados de un experimento, se propone un modelo para la variable de respuesta que tiene en cuenta las fuentes conocidas de variación como los tratamientos y, en algunos casos, las características de las unidades experimentales.

La estimación de la magnitud de la contribución de estas fuentes de variación es un objetivo principal del análisis estadístico.

Un modelo simple para un experimento con *a* tratamientos, repetidos *n* veces sobre unidades experimentales homogéneas es el siguiente:

$$Y_{ij} = \mu + \tau_i + \varepsilon_{ij}; \qquad \text{con } i=1,\ldots,a;\ j=1,\ldots,n_i$$

En este modelo Y_{ij} es la respuesta observada en la unidad experimental j-ésima del tratamiento i-ésimo, μ es una media general, τ_i es el efecto del i-ésimo tratamiento, $\mu+\tau_i$ es la respuesta esperada para el tratamiento i-ésimo y ε_{ij} es la discrepancia entre lo efectivamente observado en la unidad experimental j-ésima del tratamiento i-ésimo y la respuesta esperada para ese tratamiento. Estas discrepancias, conocidas como errores, se consideran aleatorias con esperanza cero, varianza σ^2 y estadísticamente independientes y en los modelos clásicos se suponen además normalmente distribuidas.

La evaluación de μ, τ_i y σ^2 (los llamados parámetros del modelo) es una parte central del análisis estadístico y es lo que se conoce como estimación del modelo.

Una vez que los **parámetros** se han estimado, el problema estadístico consiste en establecer si las hipótesis formuladas sobre ellos son consistentes con los resultados experimentales. La hipótesis usual es que los τ_i son iguales y equivale a la hipótesis de igualdad de los efectos de tratamientos.

Fuentes de Error

Las discrepancias entre los observado y lo esperado para un tratamiento surgen de dos fuentes principales de variación cuya magnitud relativa depende del experimento que se esté considerando. La primera es el error que se introduce cuando se quiere reproducir (repetir) el experimento sobre cada una de las unidades experimentales; la otra es la respuesta diferencial de cada unidad experimental al tratamiento que recibe y que depende de propiedades inherentes a la unidad experimental. La primera fuente de error se la conoce como **error de tratamiento** y a la segunda como **error de muestreo**.

Una vez realizado un experimento, ambas fuentes de error son indistinguibles y conforman un único error que se designa genéricamente como **error experimental.**

Existen dos recursos básicos para reducir el efecto no deseado de la presencia de los errores. Estos recursos son la aleatorización y la repetición.

Aleatorización

La aleatorización consiste en la asignación aleatoria de las unidades experimentales a los distintos tratamientos. Esta técnica tiene por objeto evitar que unidades experimentales que responden de manera particular a los tratamientos (poca respuesta, respuesta exagerada) no sean asignadas a un mismo tratamiento sino distribuidas lo mas equitativamente posible entre ellos. Lo paradójico es que la distribución equitativa, es decir, aquella que hace que las respuestas exageradas se compensen con las respuestas pobres, no es posible ya que las respuestas diferenciales no se conocen a priori y tampoco es posible saber a posteriori qué es efecto puro de tratamiento y qué es la respuesta diferencial de la unidad experimental. Sin embargo, al asignar al azar las unidades experimentales a los tratamientos, lo que se obtiene es un procedimiento que en promedio logra una distribución equitativa.

Repetición

El objetivo principal de esta técnica es lograr que la recreación del tratamiento para cada unidad experimental evite la introducción de un error sistemático en todas las unidades experimentales de un mismo tratamiento.

Este concepto es muy importante y existe una gran confusión sobre el mismo. Por ejemplo, si un tratamiento consiste en la elaboración una sustancia para posteriormente aplicarla a un lote de semillas, muchos investigadores piensan que repetir este experimento es volver a aplicar la sustancia elaborada a otro conjunto de semillas. Sin embargo, una genuina repetición es aquella que comienza con la elaboración de la sustancia desde sus componentes básicos repitiendo todas las etapas de síntesis. En muchos casos esto puede ser exagerado y para evitar trabajo innecesario el investigador debería establecer donde comienza el experimento. Para responder a esta cuestión, lo que se debe identificar es cuál es la etapa de la implementación del experimento en la que se introduce mayor variabilidad. Si esta etapa es identificada, entonces, los tratamientos deben repetirse a partir de ella.

Las repeticiones, asimismo, hacen viable el concepto de aleatorización ya que si no existen repeticiones, los efectos de tratamientos quedan **confundidos** con los efectos de parcela o unidad experimental.

En algunas áreas de conocimiento es frecuente confundir error experimental con **error de medición**. Frecuentemente el error de medición queda confundido con el error experimental, excepto cuando alícuotas o partes de una misma (o a veces única) unidad experimental son tomadas como repeticiones. En este caso se tienen **pseudo-repeticiones** y el error que se incluye en el modelo solo da cuenta del error de medición que puede ser mucho menor que el error experimental y conducir a la peligrosa creencia de que el experimento analizado es muy preciso. Más aún, cuando las pseudo-repeticiones provienen de una única unidad experimental se tendrá un experimento que puede suponerse extremadamente preciso y que, además, conduce a una estimación sesgada de los efectos de tratamiento por confusión con error de muestreo.

Precisión

Cuando un experimento es infinitamente preciso es capaz de detectar cualquier diferencia entre medias de tratamientos. Este caso ideal se obtiene cuando la varianza del error es cero, pero esta situación no ocurre en la naturaleza. Por el contrario, todos los experimentos tienen un umbral por debajo del cual no son capaces de distinguir entre tratamientos diferentes. Cuanto más preciso es el experimento más bajo es el umbral y viceversa. Por lo tanto, un objetivo principal del diseño es aumentar la precisión de un experimento. Los recursos para lograr un aumento de precisión son el incremento del número de repeticiones, el reconocimiento de fuentes sistemáticas de variación entre parcelas y, en algunos casos, el uso de experimentos factoriales.

Estructura de parcelas

Anteriormente se estableció que la aleatorización era un método de distribución equitativa de parcelas sobre y sub respondedoras a los tratamientos y que el método se justificaba en el hecho de que no era posible anticipar estas respuestas. A estos diseños en los que la aleatorización no está restringida, se los llama **completamente aleatorizados**. En algunos casos, sin embargo, es posible establecer que algunas parcelas o unidades experimentales responderán de una manera y otras de otra. Un ejemplo simple se observa en los ensayos de rendimiento cuando el terreno donde se realiza el experimento tiene una pendiente marcada. En estos casos las parcelas de la parte elevada suelen tener rendimientos menores que las de la parte baja y usar aleatorización (no restringida) como criterio de distribución de las parcelas no es la mejor decisión a la hora de planificar el experimento. Por el contrario, si a cada tratamiento se le asigna una repetición dentro de conjuntos de parcelas ubicados por ejemplo en la parte superior, media e inferior del lote experimental y se aplica aleatorización dentro de cada conjunto de esas parcelas, se habrá reconocido desde el punto de vista del diseño, una fuente sistemática de variación debida a la pendiente del terreno. Para ser consistentes con el diseño, el modelo del experimento deberá incorporar los parámetros necesarios para dar cuenta de la estructura de parcelas. El resultado de esta acción no es solo tener un modelo con más parámetros sino un experimento más preciso.

Definición 10.2: Diseño de la estructura de parcelas

El **diseño de la estructura de parcelas** consiste en el agrupamiento de unidades experimentales homogéneas en grupos o bloques.

El reconocimiento de la estructura de parcelas y su incorporación al modelo de análisis de la varianza tiene como consecuencia inmediata el aumento de precisión del diseño. Esto es así siempre y cuando la estructura de parcela obedezca al reconocimiento de variaciones reales entre las unidades experimentales ya que la imposición de una estructura de parcela arbitraria e innecesaria lejos de aumentar la precisión la disminuirá.

Un comentario final es que, si el investigador cuenta con unidades experimentales que responden homogéneamente a cada tratamiento, en cantidad suficiente para montar el experimento completo, el diseño completamente aleatorizado es preferible ya que su aplicación no requiere restricciones a la aleatorización y por lo tanto no es necesaria la estimación de parámetros adicionales. Además, en los diseños con estructura de parcela, ésta no debe interactuar con los tratamientos, *i.e.* sus efectos deben ser aditivos. En el caso del diseño completamente aleatorizado, al no existir una estructura no es necesario validar este supuesto. Finalmente, la pérdida de parcelas por diversos motivos extrínsecos a los tratamientos en los diseños completamente aleatorizados, no conduce a la aplicación de correcciones de compromiso sobre los resultados experimentales o la perdida completa de una o más repeticiones de todos los tratamientos como puede ocurrir en algunas estructuras de parcela.

Algunos diseños clásicos

A continuación se presentan tres diseños (estructura de parcelas) clásicos en la literatura de diseño de experimentos. El segundo de ellos es uno de los más simples arreglos de unidades experimentales no homogéneas y posiblemente el más popular entre los experimentadores agrícolas.

Completamente aleatorizado

Cuando las parcelas experimentales son homogéneas o no se es capaz de anticipar respuestas diferenciales de cada una de ellas, la mejor opción desde el punto de vista del diseño de experimentos es asignar los tratamientos, de manera completamente al azar. El modelo para este diseño y el análisis de la varianza discutidos en el Capítulo 8 corresponden al análisis de un experimento unifactorial sin estructura de parcelas.

Bloques completos aleatorizados

Aunque la asignación aleatoria de tratamientos es una forma natural de *distribuir* imparcialmente las pequeñas (o grandes) diferencias en las respuestas de las unidades experimentales, esta asignación no siempre es la más conveniente. Cuando las diferencias de respuestas de las unidades experimentales pueden ser anticipadas, lo conveniente es agrupar aquellas unidades similares en bloques y asignar aleatoriamente los tratamientos dentro de esos *bloques*. De esta manera, cada bloque representa una repetición completa de todos los tratamientos. Este arreglo experimental se denomina *diseño en bloques completos aleatorizados*. Se dice que son completos porque en cada bloque aparecen todos los tratamientos, y aleatorizados porque dentro de cada bloque los tratamientos son distribuidos aleatoriamente. Un caso particular de diseño en bloques es el que aparece relacionado con la prueba T para muestras apareadas, aunque el número de tratamientos es sólo dos.

Ejemplo 10.1

Se realizó un ensayo para evaluar el rendimiento en kg de materia seca por hectárea de una forrajera megatérmica con distintos aportes de N_2 en forma de urea. Las dosis de urea probadas fueron 0 (control), 75, 150, 225 y 300 kg/ha. El ensayo se realizó en distintas zonas, en las que por razones edáficas y climáticas se podían prever rendimientos diferentes. Las zonas en este caso actuaron como bloques. El diseño a campo se ilustra en la siguiente figura y a continuación se presentan los resultados obtenidos ordenados por tratamiento y por bloque.

Bloque I	225	300	75	0	150
Bloque II	300	150	75	0	225
Bloque III	75	0	300	225	150
Bloque IV	225	150	75	300	0

Figura 10.1: Asignación de tratamientos en un diseño en bloques completos aleatorizados.

Tabla 10.1: Rendimiento de materia seca (Kg/Ha) de una forrajera megatérmica con distintos niveles de aportes de nitrógeno en forma de urea.

Urea (Kg/Ha)	Bloque I	Bloque II	Bloque III	Bloque IV
0 (control)	2010	1832	2170	1879
75	2915	2175	2610	2294
150	3049	2908	2964	2971
225	3199	3235	3003	2937
300	3381	3270	3129	3171

Los datos de la tabla anterior se pueden representar genéricamente de la siguiente manera:

Tabla 10.2: Estructura típica de una tabla de datos para un ensayo unifactorial con diseño en bloques completos aleatorizados.

Tratamientos	bloque 1	bloque 2	...	bloque b	Total
1	y_{11}	y_{12}	...	y_{1b}	$y_{1\bullet}$
2	y_{21}	y_{22}	...	y_{2b}	$y_{2\bullet}$
:	:	:	:::	:	
a	y_{a1}	y_{a2}	...	y_{ab}	$y_{a\bullet}$
Total	$y_{\bullet 1}$	$y_{\bullet 2}$		$y_{\bullet b}$	$y_{\bullet\bullet}$

Cada entrada a la tabla representa una observación en el i-ésimo tratamiento (i=1,...,a) del j-ésimo bloque (j=1,...,b).

El modelo lineal para un análisis de la varianza con un factor (en este caso fertilizante) en un diseño en bloques completos, es el siguiente.

$$Y_{ij} = \mu + \tau_i + \beta_j + \varepsilon_{ij}\,, \qquad \text{con } i=1,...,a; \quad j=1,...,b$$

donde μ corresponde a la media general, τ_i el efecto del i-ésimo tratamiento, βj el efecto del j-ésimo bloque y ε_{ij} representan, como siempre, errores normales e independientes con esperanza cero y varianza común σ^2.

Respecto del modelo lineal original sólo se ha agregado el término β_j. Este término puede modelar un efecto fijo o aleatorio y este último caso supone con distribución normal independiente, esperanza cero y varianza σ^2_β e independiente del término de error. Este término modela la variación introducida por los bloques y tiene por objeto reducir el error experimental. ¿Cómo cambia la tabla de análisis de la varianza para este diseño? A continuación se muestra la Tabla de ANAVA modificada para incluir el efecto de los bloques. Calculando las cantidades para el Ejemplo 10.1, se obtienen los resultados de la Tabla 10.4.

Tabla 10.3: Fórmulas de trabajo de análisis de la varianza de un experimento unifactorial con diseño en bloques completos aleatorizados.

Fuente de Variación	Suma de Cuadrados	Grados de Libertad	Cuadrado Medio	F
Bloques	$SCB=\sum_{j=1}^{a}\frac{(y_{\bullet j})^2}{a}-\frac{(y_{\bullet\bullet})^2}{ab}$	glb=b-1		
Entre Tratamientos	$SCE=\sum_{i=1}^{b}\frac{(y_{i\bullet})^2}{b}-\frac{(y_{\bullet\bullet})^2}{ab}$	gle=a-1	$CME=\frac{SCE}{gle}$	$\frac{CME}{CMD}$
Dentro (Error Experimental)	SCD=SCT-SCE-SCB	gld=(a-1)(b-1)	$CMD=\frac{SCD}{gld}$	
Total	$SCT=\sum_{i=1}^{a}\sum_{j=1}^{b} y_{ij}^{\,2}-\frac{(y_{\bullet\bullet})^2}{ab}$	glt=ab-1		

Tabla 10.4: Tabla de análisis de la varianza para el rendimiento de materia seca (Kg/Ha) de una forrajera megatérmica con distintos aportes de N_2 en forma de urea.

Fuente de Variación	Suma de Cuadrados	Grados de Libertad	Cuadrado Medio	F
Bloques	203319.0	3	67773.0	
Entre Tratamientos	4291440.0	4	1072860.0	41.57
Dentro	309716.5	12	25809.7	
Total	4804475.5	19		

El procedimiento del test de hipótesis es similar al realizado para un diseño completamente aleatorizado. Dado que F, 41.57, es mayor que el cuantil (1-α) de una distribución $F_{4,12}$ se rechaza la hipótesis de igualdad de tratamientos. La aplicación del test a posteriori es directa y el número de bloques (b) sustituye el número de repeticiones en el cálculo del error estándar de la comparación.

Cuadrado latino

Una extensión directa del concepto de bloques completos aleatorizado es la del cuadrado latino, en el que se incorporan al diseño, el reconocimiento de dos fuentes sistemáticas de variación entre parcelas.

Este diseño no es tan popular como el anterior ya que impone un número fijo de repeticiones y cuando el número de tratamientos es grande, el experimento completo puede ser inmanejable. De hecho, el número total de parcelas experimentales es igual al cuadrado del número de tratamientos. No obstante estas dificultades, el cuadrado latino es un diseño base de otros diseños como los llamados experimentos cross-over, populares en la experimentación con animales.

El diseño en cuadrado latino clásico de la experimentación agrícola, en el que ensayan a tratamientos, se obtiene ordenando a^2 parcelas experimentales en un cuadrado de $a \cdot a$ parcelas y asignando a parcelas a cada uno de los tratamientos de tal manera que en cada fila y en cada columna haya sólo una repetición de cada tratamiento como muestra la Figura 10.2.

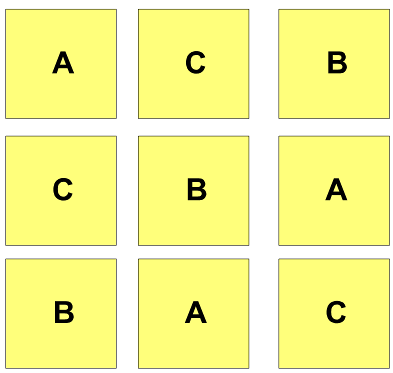


Figura 10.2: Diseño en cuadrado latino para un experimento en el que se ensayan tres tratamientos (A,B y C).

El modelo lineal de un experimento en diseño cuadrado latino es el siguiente:

$$Y_{ijk} = \mu + \tau_i + \chi_j + \rho_k + \varepsilon_{ijk} \quad \text{con } i=1,\ldots,a;\ j=1,\ldots,a;\ k=1,\ldots,a$$

donde Y_{ijk} es la observación de la respuesta del i-ésimo tratamiento en la columna j-ésima y fila k-ésima. ε_{ijk} es el término de error correspondiente a la observación del i-ésimo tratamiento en la columna j-ésima y fila k-ésima. En este modelo los parámetros χ_j y ρ_k modelan los efectos de las columnas y las filas respectivamente.

El cuadro de Análisis de la Varianza para este diseño se calcula según las expresiones provistas en la Tabla 10.5.

Tabla 10.5 :Fórmulas de trabajo de análisis de la varianza de un experimento unifactorial con diseño en Cuadrado Latino.

Fuente de Variación	Suma de Cuadrados	Grados de Libertad	Cuadrado Medio	F
Filas	$SCF=\sum_{k=1}^{a}\frac{(y_{i\bullet k})^2}{a}-\frac{(y_{\bullet\bullet\bullet})^2}{a\cdot a}$	a-1		
Columnas	$SCC=\sum_{j=1}^{a}\frac{(y_{ij\bullet})^2}{a}-\frac{(y_{\bullet\bullet\bullet})^2}{a\cdot a}$	a-1		
Entre Tratamientos	$SCE=\sum_{i=1}^{a}\frac{(y_{i\bullet\bullet})^2}{a}-\frac{(y_{\bullet\bullet\bullet})^2}{a\cdot a}$	gle=a-1	$CME=\frac{SCE}{gle}$	$\frac{CME}{CMD}$
Dentro (Error Experimental)	SCD=SCT-SCF-SCC-SCE	gld=(a-1)(a-2)	$CMD=\frac{SCD}{gld}$	
Total	$SCT=\sum_{i=1}^{a}\sum_{j;k=1}^{a} y_{ijk}^{2}-\frac{(y_{\bullet\bullet\bullet})^2}{a\cdot a}$	$glt=a^2-1$		

Ejemplo 10.2

La siguiente tabla muestra los rendimientos de remolacha azucarera en toneladas por hectárea bajo tres tipos de labores culturales.

Tabla 10.6: Rendimiento de remolacha azucarera en toneladas por hectárea bajo tres tipos de labores culturales obtenidos de un experimento en cuadrado latino.

	Col I	Col II	Col III
Fila I	130 (A)	90 (C)	140 (B)
Fila II	100 (C)	120 (B)	147 (A)
Fila III	133 (B)	125 (A)	115 (C)

La Tabla 10.7 presenta el cuadro de análisis de la varianza correspondiente que muestra un efecto significativo de los distintos métodos culturales aplicados. Es importante notar que la suma de cuadrados debida a las columnas es muy importante y si no hubiera sido removida de la suma de cuadrados del error la interpretación de estos resultados hubiera sido diferente.

Tabla 10.7: Análisis de la varianza para el experimento de rendimiento de remolacha azucarera.

Fuente de Variación	Suma de Cuadrados	Grados de Libertad	Cuadrado Medio	F
Filas	28.2	2		
Columnas	754.9	2		
Entre Tratamientos	1914.9	2	957.4	344.7
Dentro (Error Experimental)	5.5	2	2.8	
Total	2703.5	8		

Estructura de tratamientos

En un punto anterior se presentó a los tratamientos como los distintos niveles de un único factor o como combinación de niveles de varios factores. En este último caso, el experimentador se pregunta si es posible identificar los efectos de cada uno de los factores, estimarlos y eventualmente probar hipótesis sobre ellos. Aunque la respuesta es afirmativa aún persiste una duda fundamental ¿para qué diseñar experimentos en los que hay que usar herramientas analíticas especiales para separar los efectos de los distintos factores si se pueden planificar experimentos más sencillos para cada factor evitando complicaciones? La respuesta a este problema está relacionada con el concepto de eficiencia y que en términos prácticos se relaciona con la cantidad de repeticiones que son necesarias en un experimento para tener una precisión dada. Por ejemplo si para evaluar los efectos de los factores A y B con tres niveles cada uno se requieren tres repeticiones para cada nivel, se necesitarán 9 unidades experimentales para el ensayo del factor A y otras 9 para el ensayo del factor B, haciendo un total de 18 unidades experimentales. Si en vez de utilizar dos experimentos separados se planifica un experimento conjunto con 9 tratamientos (3 niveles de A x 3 niveles de B) y solo se repite una vez cada tratamiento, solo se necesitarán 9 unidades experimentales para acomodar todo el experimento y aún se tendrán tres unidades tratadas con cada uno de los niveles de cada uno de los factores. Es decir que, aunque no se cuentan con repeticiones para las combinaciones de niveles de factores, si las hay (tres) para cada uno de los niveles de los factores individuales. En consecuencia, con la mitad de las unidades experimentales necesarias para acomodar los experimentos separados, se puede montar un experimento conjunto que provee la misma precisión para la evaluación de cada factor individual. Si aún se quisieran invertir las 18 unidades experimentales de los dos experimentos originales, se podría hacer una repetición completa de todo el experimento y se tendría el doble de unidades experimentales para cada nivel de cada uno de los factores y en este sentido, los experimentos factoriales son más eficientes para evaluar los efectos de los factores individuales. Pero los experimentos factoriales, cuando están repetidos, permiten además, probar la existencia y estimar la magnitud de respuestas diferenciales a la combinación de los factores individuales, fenómeno que se conoce como interacción. Dado que la interacción es común en los sistemas biológicos, los experimentos que son capaces de detectarla y estimarla son siempre preferibles.

Definición 10.3: Estructura de Tratamientos

La **estructura de tratamientos** de un diseño de experimentos consiste en el conjunto de tratamientos o poblaciones que el experimentador ha seleccionado para estudiar y/o comparar.

Experimentos Factoriales

En los modelos de los experimentos factoriales los parámetros τ_i que hacen referencia a los efectos de tratamientos se descompone en un conjunto de parámetros que dan cuenta de cada uno de los factores intervinientes y se agrega según sea necesario, conveniente y posible, los términos correspondientes a las interacciones.

Modelos aditivos

Los modelos factoriales **aditivos** son aquellos en los que los términos que modelan la interacción están ausentes. Para ejemplificar este caso se presenta un experimento factorial 2x2 (dos factores con dos niveles cada uno) en el que la interacción se supone ausente y montado en un diseño completamente aleatorizado. Los Factores se han designado como A y B y sus niveles como A_1,A_2 y B_1,B_2. Como existen 4 tratamientos (A_1B_1, A_1B_2, A_2B_1, A_2B_2) y estos no están repetidos, se necesitan sólo cuatro parcelas experimentales. Dado que el diseño es completamente aleatorizado la asignación de las parcelas a cada uno de los tratamientos es al azar. Un arreglo posible se presenta en la siguiente figura.

A_2B_1	A_1B_2
A_1B_1	A_2B_2

Figura 10.3: Experimento bifactorial sin repeticiones montado en un diseño completamente aleatorizado.

El modelo para este experimento es el siguiente:

$$Y_{ij}=\mu+\alpha_i+\beta_j+\varepsilon_{ij} \quad \text{con } i=1,2;\ j=1,2$$

En este modelo Y_{ij} representa la respuesta al i-ésimo nivel del factor A y j-ésimo nivel de factor B, μ representa una media general, α_i el efecto que produce el i-ésimo nivel del factor A, β_j corresponde al j-ésimo nivel del factor B y ε_{ij} es el error asociado a la observación ij-ésima que como siempre se suponen normales, independientes, con esperanza cero y varianza común σ^2.

El cuadro de Análisis de la Varianza para este diseño se calcula según las expresiones provistas en la Tabla 10.8.

Tabla 10.8:Expresiones para el cálculo del cuadro de análisis de la varianza de un experimento bifactorial con diseño completamente aleatorizado.

Fuente de Variación	Suma de Cuadrados	Grados de Libertad	Cuadrado Medio	F
Factor A	$SCF=\sum_{i=1}^{a}\frac{(y_{i\bullet})^2}{b}-\frac{(y_{\bullet\bullet})^2}{a \cdot b}$	gla=a-1	$CMA=\frac{SCA}{gla}$	$\frac{CMA}{CMD}$
Factor B	$SCC=\sum_{j=1}^{b}\frac{(y_{\bullet j})^2}{a}-\frac{(y_{\bullet\bullet})^2}{a \cdot b}$	glb=b-1	$CMB=\frac{SCB}{glb}$	$\frac{CMB}{CMD}$
Dentro (Error Experimental)	SCD=SCT-SCA-SCB	gld=(a-1)(b-1)	$CMD=\frac{SCD}{gld}$	
Total	$SCT=\sum_{i=1}^{a}\sum_{j=1}^{b} y_{ij}^2-\frac{(y_{\bullet\bullet})^2}{a \cdot b}$	glt=a·b-1		

Ejemplo 10.3

En un ensayo comparativo del efecto del estrés hídrico y salino sobre la germinación de *Atriplex cordobensis*, se sometieron lotes de semillas a cuatro niveles de potencial agua: 0, -0.5, -1.0 y –1.5 Mpa obtenidos mediante la aplicación al medio de dos osmolitos: polietilenglicol (PEG) o cloruro de sodio (ClNa). El experimento se montó en un diseño completamente aleatorizado sin repeticiones cuyos resultados se presentan en la siguiente tabla.

Tabla 10.9: Resultados de un ensayo comparativo del efecto de distintos potenciales agua del substrato obtenido con dos osmolitos: polietilenglicol(PEG) y cloruro de sodio (ClNa) sobre el porcentaje de germinación en A. cordobensis.

Mpa	*0*	*-0.5*	*-1.0*	*-1.5*
ClNa	85	78	54	14
PEG	83	76	43	9

Cuando los experimentos factoriales no tienen repeticiones, el analista debe suponer que los factores no interactúan para poder estimar la varianza del error experimental. Si este supuesto no se cumple entonces el experimento está deficientemente diseñado y las conclusiones del análisis pueden ser completamente erróneas. Existen algunas pruebas para verificar este supuesto como la prueba de aditividad de Tukey (1949).

La tabla de análisis de la varianza para este experimento, suponiendo un modelo aditivo, se muestra en la siguiente tabla.

Tabla 10.10: Cuadro de análisis de la varianza para de un experimento bifactorial para evaluar el efecto de distintos potenciales agua del substrato obtenidos por el agregado al medio de dos osmolitos: polietilenglicol(PEG) o cloruro de sodio (ClNa) sobre el porcentaje de germinación en A. cordobensis.

Fuente de Variación	Suma de Cuadrados	Grados de Libertad	Cuadrado Medio	F
Osmolito	50.0	1	50.0	5.6
Potencial Agua	6118.5	3	2172.8	241.4
Dentro	27.0	3	9.0	
Total	6195.5	7		

Consultando los valores críticos de una F con 1 y 3 grados de libertad para el factor osmolito y con 3 y 3 grados de libertad para potencial agua, se puede apreciar que ambos factores afectan significativamente el porcentaje de germinación.

Modelos con interacción

Si el experimentador supone o sospecha que la respuesta a dos o más factores no se puede explicar como la suma de sus efectos individuales entonces el modelo para el experimento factorial deberá incluir términos de interacción que den cuenta de este hecho. La inclusión de términos de interacción en el modelo conlleva la necesidad de tener repeticiones para cada tratamiento porque de otra forma no es posible estimar los parámetros adicionales. Aunque no se profundizará más en este tema, cuando el experimento tiene dos factores, existen solo interacciones de primer orden, cuando tiene tres factores, existen interacciones de primer y de segundo orden y así sucesivamente para factoriales de mayor orden.

A continuación se examinará con algún nivel de detalle un experimento bifactorial con interacción y se presentará un ejemplo.

El modelo para un experimento bifactorial con interacciones es una ampliación del modelo para el experimento bifactorial descripto anteriormente, excepto que incluye un conjunto adicional de parámetros, conocidos como de interacción.

$$Y_{ijk}=\mu+\alpha_i+\beta_j+\delta_{ij}+\varepsilon_{ijk} \quad \text{con } i=1,2;\ j=1,2;\ k=1,..,n_{ij}$$

En este modelo Y_{ijk} representa la respuesta de la k-ésima repetición en el i-ésimo nivel del factor A y j-ésimo nivel de factor B, μ representa una media general, α_i el efecto que produce el i-ésimo nivel del factor A, β_j corresponde al efecto del j-ésimo nivel del factor B, δ_{ij} los efectos adicionales (interacciones) para cada combinación de los niveles de los factores y ε_{ijk} es el error asociado a la observación ijk-ésima que como siempre se supone normal e independiente con esperanza cero y varianza común σ^2. Debe notarse que el subíndice *k* se mueve entre 1 y n_{ij}, es decir, el número de repeticiones para el tratamiento puede ser distinto.

El cuadro de Análisis de la Varianza para este diseño se calcula según las expresiones provistas en la Tabla 10.11.

Tabla 10.11: Expresiones para el cálculo del cuadro de análisis de la varianza de un experimento bifactorial con interacción en un diseño completamente aleatorizado.

Fuente de Variación	Suma de Cuadrados	Grados de Libertad	Cuadrado Medio	F
Factor A	$SCF=\sum_{i=1}^{a}\frac{(y_{i\bullet\bullet})^2}{n_{i\bullet}}-\frac{(y_{\bullet\bullet\bullet})^2}{n_{\bullet\bullet}}$	gla=a-1	$CMA=\frac{SCA}{gla}$	$\frac{CMA}{CMD}$
Factor B	$SCC=\sum_{j=1}^{b}\frac{(y_{\bullet j\bullet})^2}{n_{\bullet j}}-\frac{(y_{\bullet\bullet\bullet})^2}{n_{\bullet\bullet}}$	glb=b-1	$CMB=\frac{SCB}{glb}$	$\frac{CMB}{CMD}$
Interacción AB	$SCAB\sum_{i=1}^{a}\sum_{j=1}^{b}\frac{(y_{ij\bullet})^2}{n_{ij}}-\frac{(y_{\bullet\bullet\bullet})^2}{n_{\bullet\bullet}}$	Glab= (a-1)(b-1)	$CMAB=\frac{SCAB}{glab}$	$\frac{CMAB}{CMD}$
Dentro (Error Experimental)	SCD=SCT-SCA-SCB-SCAB	gld=glt-gla-glb-glab	$CMD=\frac{SCD}{gld}$	
Total	$SCT=\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n_{ij}}y_{ijk}^2-\frac{(y_{\bullet\bullet\bullet})^2}{n_{\bullet\bullet}}$	glt= $n_{\bullet\bullet}$-1		

Ejemplo 10.4

En un estudio sobre la potencialidad forrajera de *Atriplex cordobensis*, un arbusto que crece en depresiones del chaco árido argentino, se evaluó la concentración de proteínas en hojas cosechadas en invierno y verano sobre plantas masculinas y femeninas. Para cada combinación de sexo y estación, se obtuvieron tres determinaciones del contenido proteico medido como porcentaje del peso seco. Los resultados se presentan en la siguiente tabla.

Tabla 10.12: Concentración proteica (% del peso seco) en hojas de Atriplex cordobensis cosechadas en invierno y verano de plantas masculinas y femeninas.

Femeninas		Masculinas	
Invierno	Verano	Invierno	Verano
24	17	17	24
28	18	18	25
26	16	16	23

La tabla que presenta los resultados del análisis de la varianza se muestra a continuación. Como puede observarse, ninguno de los factores ensayados muestra por si mismo un efecto significativo sobre la concentración de proteínas pero el término de interacción es altamente significativo, indicando que los factores estudiados efectivamente intervienen en la expresión final de la concentración de proteínas pero que sus efectos no son independientes del nivel del otro factor. La Figura 10.4 presenta una representación gráfica de los valores medios en los cuatro tratamientos que permite interpretar fácilmente el resultado mostrado en el cuadro de análisis de la varianza.

Tabla 10.13: Cuadro de análisis de la varianza para el efecto del sexo y la época de cosecha sobre la concentración de proteínas en hojas de Atriplex cordobensis.

Fuente de Variación	Suma de Cuadrados	Grados de Libertad	Cuadrado Medio	F
Factor Sexo	3	1	3.00	1.71
Factor Época de cosecha	3	1	3.00	1.71
Interacción Época-Sexo	192	1	192.00	109.71
Dentro	14	8	1.75	
Total	212	11		

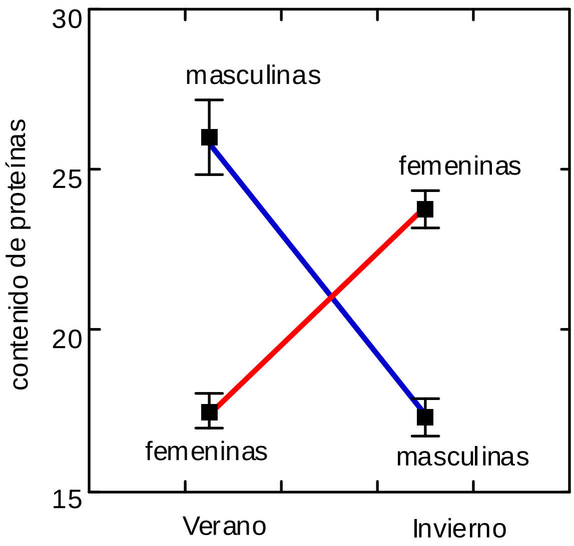


Figura 10.4: Media ± error estándar de la concentración de proteínas en hojas de Atriplex cordobensis por efecto del sexo y la época de cosecha.

Los modelos con interacción no siempre muestran comportamientos tan extremos como el del ejemplo anterior. De hecho, en muchas situaciones los *perfiles de respuesta* no se cruzan aunque exista interacción significativa. Un ejemplo se puede apreciar en la Figura 10.5 que también corresponde al trabajo sobre potencialidad forrajera de *A. cordobensis*, pero en este caso la variable estudiada es la proporción de fibras insolubles. Las plantas masculinas siempre presentaron mayor contenido de fibras insolubles que las femeninas (efecto principal del factor sexo) pero la diferencia entre femeninas y masculinas fue mayor en el invierno que en el verano (interacción). De igual modo se puede interpretar el efecto de la época de cosecha, diciendo que en verano el contenido de fibras insolubles fue siempre mayor que en invierno (efecto principal del factor época de cosecha) pero que esta diferencia es más marcada en las plantas femeninas.

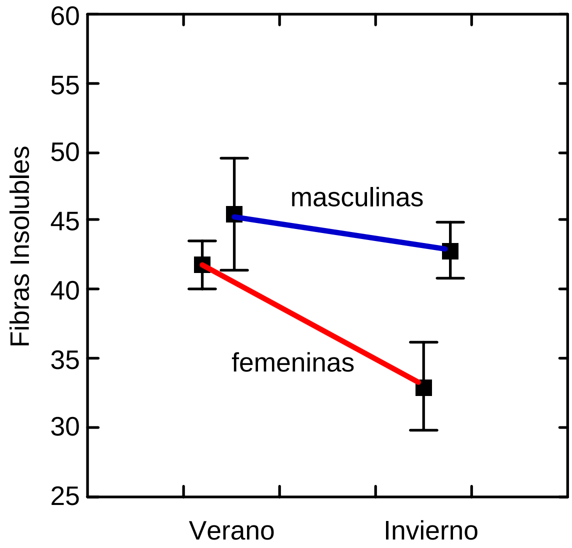


Figura 10.5: Media ± error estándar de la concentración de fibras insolubles en hojas de Atriplex cordobensis por efecto del sexo y la época de cosecha.

Aunque en los ejemplos anteriores se han presentado experimentos con estructura factorial de tratamientos sólo en diseños completos al azar, la combinación de estructuras factoriales y estructuras de parcela da lugar a una amplia variedad de arreglos experimentales. Así, un experimento como aquel en que se evaluaba el efecto de dos osmolitos y cuatros potenciales agua sobre la germinación se podría haber diseñado con repeticiones. Por diversas razones, quizás no puedan asegurarse las mismas condiciones experimentales de repetición a repetición, porque, por ejemplo, no siempre las cámaras de cultivo regulan de manera similar la temperatura del ensayo, y por lo tanto se tiene una fuente potencial de variación conocida que no es de interés por si misma pero que sí debe incorporarse al diseño y al modelo para eliminarla del error experimental. De este modo, cada repetición podría considerarse un bloque y el experimento completo sería un experimento con estructura bifactorial de tratamiento y estructura de parcelas en bloques completos al azar.

Como se anticipó al comienzo del capítulo, los temas de diseño experimental no se agotan en esta presentación. Queda una importante variedad de tópicos relativos a jerarquías en la estructura de parcelas y en la estructura de tratamientos, métodos de partición de sumas de cuadrados, análisis de la interacción en los modelos no aditivos, análisis de la covarianza, modelos con factores de efectos aleatorios, modelos con mezcla de factores con efectos aleatorios y fijos, diseño del número de repeticiones para alcanzar una potencia deseada, etc, etc, que el lector interesado deberá consultar en una obra más completa de Diseño de Experimentos (Montgomery, 1991; Milliken and Johnson, 1995; Steel y Torrie, 1985).

Parcelas Divididas

Este diseño esta asociado a los experimentos factoriales. Por ejemplo, si los tratamientos están conformados por dos factores, el diseño en parcelas divididas tiene la particularidad de asociar los niveles de un factor a parcelas o unidades experimentales grandes (*parcelas principales*) y los niveles del otro factor a *subparcelas* obtenidas por división de las parcelas principales. Las parcelas principales pueden no mostrar estructura y el factor principal asignarse al azar a un número *r* de parcelas principales (diseño de parcelas divididas completamente aleatorizado) o bien las parcelas principales pueden estructurarse en bloques (diseño de parcelas divididas en bloques). La característica más sobresaliente de este diseño es la presencia de dos errores experimentales diferentes. Uno que representa la variabilidad entre parcelas principales tratadas de la misma forma (error de las parcelas principales) y otro que representa la variabilidad entre subparcelas tratadas de la misma forma (error de las subparcelas). El cuadrado medio asociado al error experimental de las parcelas principales se utiliza como denominador del estadístico F en los contrastes de hipótesis de los factores asociados a las parcelas principales, así como también en las pruebas de comparaciones múltiples entre las medias de los tratamientos aplicados a las parcelas principales. El cuadrado medio del error de subparcela se utiliza para el contraste de hipótesis de los factores asociados a las subparcelas y de las interacciones de éstos con el o los factores asociados a las parcelas principales así como para las correspondientes comparaciones múltiples. Una representación gráfica de un diseño en parcelas divididas en bloques se muestra en la Figura 10.6.

A modo de ejemplo, en la investigación agronomía un factor, como el método de cultivo, puede requerir equipos que sólo se pueden usar en parcelas grandes (parcela principal); mientras que otros factores, como la fertilización, con sus distintas variantes, se podrían aplicar a una parcela más pequeña (subparcela). En el área de investigación educativa se pueden aplicar diferentes métodos de enseñanza a los alumnos de distintas divisiones (parcelas principales) y estudiar un factor adicional, como puede ser el uso de ciertos materiales didácticos o computadoras, aplicado a subgrupos de alumnos (subparcelas). En la investigación industrial un fabricante de papel puede probar tres métodos diferentes para preparar pulpa y cuatro diferentes métodos de cocción, evaluando la resistencia del papel. Para ello, podría producir un con cada método un lote de pulpa (parcela principal) y subdividirlo en cuatro partes (subparcelas), asignándose a cada una de ellas los distintos métodos de cocción.

El modelo lineal, para un experimento bifactoral conducido según un diseño en parcelas divididas en bloques es el siguiente:

$$Y_{ijk} = \mu + \tau_i + \beta_j + (\tau\beta)_{ij} + \gamma_{ijk} + (\tau\gamma)_{ik} + \varepsilon_{ijk}$$

Donde Y_{ijk} representa la observación en el k-ésimo nivel del factor aplicado a la subparcela, de la i-ésima parcela principal en el j-ésimo bloque, τ_i representa el i-ésimo nivel del factor aplicado a la parcela principal, β_j el j-ésimo bloque, $(\tau\beta)_{ij}$ el error experimental de las parcelas principales (variación aleatoria entre parcelas principales tratadas de la misma forma), que se simboliza como la interacción entre el factor principal y los bloques ya que su suma de cuadrados se calcula como la suma de cuadrados de esa interacción, γ_{ijk} representa el efecto del k-ésimo nivel del factor asociado a la subparcela dentro de la i-ésima parcela principal del j-ésimo bloque, $(\tau\gamma)_{ik}$ representa la interacción del factor principal con el factor aplicado a las subparcelas y ε_{ijk} el error experimental a nivel de subparcelas.

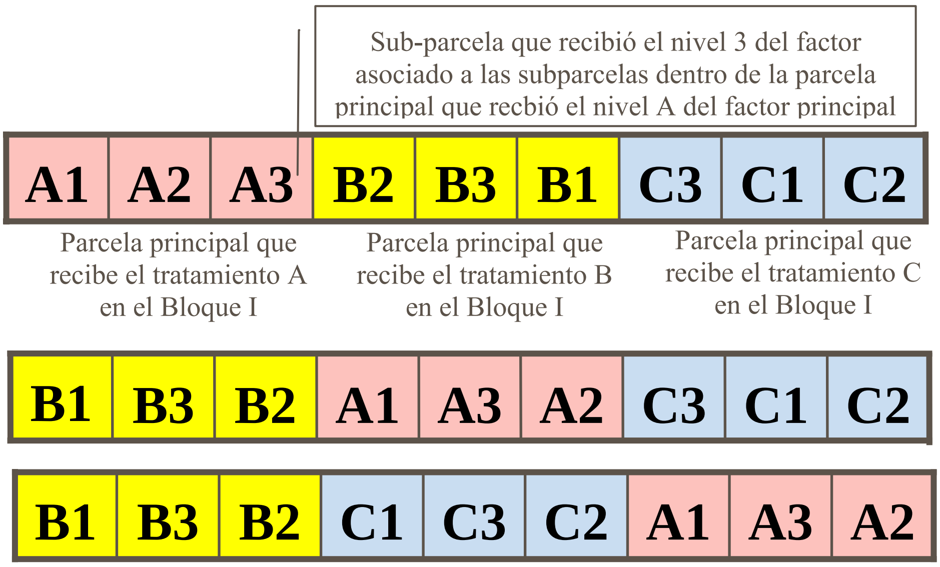


Figura 10.6: Parcelas divididas en bloques. En cada bloque se muestran los tres niveles para el factor principal (A, B y C) y los tres niveles para el factor de las subparcelas (1, 2 y 3).

La Tabla del análisis de la varianza para este diseño es la siguiente.

Tabla 10.14: Esquema de tabla de análisis de la varianza para un experimento bifactorial conducido según un diseño en parcelas divididas en bloques con "r" repeticiones. El factor asociado a las parcelas principales se designa como Factor A y tiene "a" niveles y el factor B con "b" niveles es aplicado a las subparcelas.

Fuente de Variación	Suma de Cuadrados (sc)	Grados de Libertad (gl)	Cuadrado Medio (cm)	F
Bloques (R)	scR	glR=r-1		
Factor en parcela principal (A)	scA	glA=a-1	$cmA=\frac{scA}{glA}$	$F=\frac{cmA}{cmEA}$
Error Parcela Principal	scEA	glEA=(a-1)(b-1)	$cmEA=\frac{scEA}{glEA}$	
Factor en subparcela (B)	scB	glB=(b-1)	$cmB=\frac{scB}{glB}$	$F=\frac{cmB}{cmESP}$
Interacción BxA	scAB	glAB=(a-1)(b-1)	$cmAB=\frac{scAB}{glAB}$	$F=\frac{cmAB}{cmESP}$
Error subparcela (ESP)	SCSP=SCT-scR-scA-scEA-scB-scAB	glESP=glT-glR-glA-glEA-glB-glAB	$cmESP=\frac{SCSP}{glESP}$	
Total	SCT	glT=N-1		

Ejemplo 10.5

En un ensayo de trigo se dispusieron dos parcelas principales en tres bloques. Sobre las parcelas principales se aleatorizaron los niveles del factor riego y estas fueron divididas en cuatro subparcelas donde se aleatorizaron cuatro variedades de trigo. La variable en estudio fue el rendimiento medido en kg/parcela experimental. Para el factor "riego"(Factor A) se tienen dos niveles: secano (sin riego) y riego y para el factor "variedad" (Factor B) se usaron las siguientes variedades: Buck-Charrúa (B-Ch), Las Rosas-INTA (LR-INTA), Pigue y Pro-INTA Puntal (P-INTA P). Los datos (gentileza Ing. M. Cantarero, Facultad de Ciencias Agropecuarias, U.N.C.) se presentan a continuación y se encuentran bajo el nombre *ParcelaD.idb* entre los datos de prueba disponibles en el paquete estadístico InfoStat.

Tabla 10.15: Datos de un ensayo comparativo de rendimientos de cuatro variedades de trigo bajo riego y secano, conducido en un diseño en parcelas divididas (riego-secano) repetidas en tres bloques.

Parc	**Blo**	**Variedad**	**Rend**	**Parc**	**Blo**	**Variedad**	**Rend**	**Parc**	**Blo**	**Variedad**	**Rend**
R	1	B-Ch	409.3	R	2	B-Ch	311.7	R	3	B-Ch	516.4
R	1	LR-INTA	544.9	R	2	LR-INTA	445.4	R	3	LR-INTA	585.7
R	1	Pigue	519.9	R	2	Pigue	477.0	R	3	Pigue	624.5
R	1	P-INTA P	629.5	R	2	P-INTA P	639.0	R	3	P-INTA P	585.7
S	1	B-Ch	266.3	S	2	B-Ch	252.8	S	3	B-Ch	299.9
S	1	LR-INTA	259.3	S	2	LR-INTA	358.4	S	3	LR-INTA	350.2
S	1	Pigue	340.7	S	2	Pigue	296.6	S	3	Pigue	327.2
S	1	P-INTA P	236.6	S	2	P-INTA P	335.7	S	3	P-INTA P	390.5

Tabla 10.16: ANAVA para un ensayo comparativo de rendimientos de cuatro variedades de trigo bajo riego y secano, conducido en un diseño en parcelas divididas en bloques.

Fuente de Variación	Suma de Cuadrados	Grados Libertad	Cuadrado Medio	F	p-valor
Bloque	22912.97	2	11456.48		
Riego	276233.13	1	276233.13	*55.24*	*0.0176*
Bloque x Riego (EEPP)	10001.49	2	*5000.74*		
Variedad	51095.57	3	17031.86	**6.384**	**0.0078**
Riego x Variedad	18926.16	3	6308.72	**2.364**	**0.1224**
Error Experimental (EESP)	32014.74	12	**2667.90**		
Total	411184.06	23			

En itálica se destaca el cuadrado medio utilizado para la obtención del estadístico F que aparece en itálica y en negrilla el cuadrado medio para los estadísticos F que aparecen en negrilla.

De acuerdo a los resultados, hay una diferencia estadísticamente significativa entre los rendimientos de trigo creciendo bajo riego o secano (F=55.24 con 1 y 2 gl, p=0.0176). Esta diferencia no depende de la variedad (interacción variedad por riego no significativa, F=2.364, 3 y 12 gl, p=0.1224) pero si hay diferencias entre variedades (F=6.384, 3 y 12 gl, P=0.0078).

Ejercicios

Ejercicio 10.1

El siguiente conjunto de datos corresponde a proteína bruta en leche obtenida con dos suplementos (A y B) en dos dosis (1 y 2). Cada observación corresponde al contenido de proteína bruta en leche de una muestra obtenida de una muestra amalgamada por tambo.

Tambo	Control	A1	A2	B1	B2
I	3.19	3.03	3.06	3.22	3.33
II	3.16	3.07	3.08	3.28	3.20
III	3.25	3.23	3.24	3.45	3.45
IV	3.48	3.30	3.33	3.44	3.39
V	3.25	3.25	3.24	3.35	3.54
VI	3.10	3.05	2.93	3.28	3.35

a) Calcular la estadística descriptiva básica.

b) Identificar el modelo lineal para los datos anteriores.

c) Calcular la tabla de análisis de la varianza y, si corresponde, utilizar alguna técnica de comparaciones múltiples.

d) ¿Qué suplementación se recomendaría si el objetivo es maximizar la concentración de proteína bruta en la leche?

Ejercicio 10.2

En un experimento sobre la incidencia de una virosis sobre el perímetro de las cabezas de ajo blanco, se comparó el perímetro medio de las cabezas obtenidas de plantas libre de virus y de plantas enfermas, bajo dos frecuencias de riego: cada 15 días y cada 30 días. El experimento se realizó siguiendo un diseño completamente aleatorizado con tres repeticiones donde la unidad experimental era una parcela de 3 surcos de 5 metros cada uno y de los cuales sólo se tomó el surco central para evitar efectos de bordura. Los resultados obtenidos fueron los siguientes:

Plantas Sanas		Plantas Enfermas	
Riego c/15d	Riego c/30d	Riego c/15d	Riego c/30d
45.5	40.1	41.5	35.8
43.0	37.3	37.0	31.4
41.3	38.1	36.3	33.8

a) Identificar el modelo lineal.

b) Construir la tabla de análisis de la varianza para este modelo.

c) Concluir sobre el efecto de la virosis, el riego y su eventual interacción.

Ejercicio 10.3

En la siguiente tabla se muestran los resultados de un experimento montado según un diseño completamente aleatorizado con cuatro repeticiones, en el que nemátodos de género *Pratylenchus* fueron criados en cuatro condiciones de temperatura y discriminados según sexo para evaluar el efecto del sexo y la temperatura sobre la expresión fenotípica de diversos caracteres morfométricos. Los resultados presentados corresponden al largo promedio de la cola en unidades experimentales conformadas por 5 individuos.

	Hembras				**Machos**			
Temp. (ºC)	Rep 1	Rep 2	Rep 3	Rep 4	Rep 1	Rep 2	Rep 3	Rep 4
16	29.2	32.5	34.6	32.6	27.2	24.7	27.3	26.2
21	30.1	30.4	31.4	35.8	26.7	26.5	27.2	27.2
25	31.6	30.2	29.5	30.0	26.2	26.3	28.2	26.2
28	29.6	28.4	28.4	28.1	24.8	25.4	25.6	26.2

a) Identificar el modelo lineal para este experimento.

b) Representar gráficamente los valores medios según sexo y temperatura.

c) Construir la tabla de análisis de la varianza correspondiente.

d) Concluir sobre el efecto de la temperatura y el sexo sobre la expresión del largo de la cola y relacione sus conclusiones con la representación gráfica obtenida en b).

Ejercicio 10.4

Considere el Ejercicio 10.3 y suponga que debido al tamaño del experimento las repeticiones se realizaron en laboratorios diferentes. Considere que las repeticiones como bloques.

a) Identificar el modelo lineal para las observaciones de este experimento.

b) Construir una tabla de análisis de la varianza.

c) Concluir sobre la acción del sexo, la temperatura y su eventual interacción.

Ejercicio 10.5:

Se realizó un experimento con seis variedades de lechuga en tres fechas de siembra. La variable respuesta fue el rendimiento. El ensayo condujo según un diseño de parcelas divididas en bloques con cuatro repeticiones. Las fechas de siembra estuvieron asociadas a las parcelas principales y las variedades a las subparcelas. Los datos que se presentan a continuación se encuentran bajo el nombre *tabla ejer. 10.5.IDB* en la carpeta *Unidad10* dentro de la carpeta que contiene los datos de los ejercicios del presente libro y que se encuentra en la sección de *Material didáctico* en: http://www.agro.uncor.edu/~estad/.

Fecha	Var	Blo	Rend
1	A	1	11.80
1	B	1	8.30
1	C	1	9.20
1	D	1	15.60
1	E	1	16.20
1	F	1	9.90
1	A	2	7.50
1	B	2	8.40
1	C	2	10.60
1	D	2	10.80
1	E	2	11.20
1	F	2	10.80
1	A	3	9.70
1	B	3	11.80
1	C	3	11.40
1	D	3	10.30
1	E	3	14.00
1	F	3	4.80
1	A	4	6.40
1	B	4	8.50
1	C	4	7.20
1	D	4	14.70
1	E	4	11.50
1	F	4	9.80

Fecha	Var	Blo	Rend
2	A	1	9.70
2	B	1	5.40
2	C	1	12.10
2	D	1	13.20
2	E	1	16.50
2	F	1	12.50
2	A	2	8.80
2	B	2	12.90
2	C	2	15.70
2	D	2	11.30
2	E	2	11.10
2	F	2	14.30
2	A	3	12.50
2	B	3	11.20
2	C	3	7.60
2	D	3	11.00
2	E	3	10.80
2	F	3	15.90
2	A	4	9.40
2	B	4	7.80
2	C	4	9.40
2	D	4	10.70
2	E	4	8.50
2	F	4	7.50

Fecha	Var	Blo	Rend
3	A	1	7.00
3	B	1	5.70
3	C	1	3.30
3	D	1	12.60
3	E	1	12.60
3	F	1	10.20
3	A	2	9.10
3	B	2	8.40
3	C	2	6.90
3	D	2	15.40
3	E	2	12.30
3	F	2	11.60
3	A	3	7.10
3	B	3	6.10
3	C	3	1.00
3	D	3	14.20
3	E	3	14.40
3	F	3	10.40
3	A	4	6.30
3	B	4	8.80
3	C	4	2.60
3	D	4	11.30
3	E	4	14.10
3	F	4	12.20

a) Proponer un modelo estadístico adecuado para este experimento y plantear las hipótesis a contrastar.

b) Realizar el análisis de la varianza para determinar si hay diferencias entre variedades y entre fechas.

c) Analizar la interacción entre ambos factores. ¿Recomendaría alguna variedad?

d) Analizar si se verifican los supuestos del modelo.

11

Análisis de Datos Categóricos

Introducción

En cualquier área del conocimiento, tal como la Agronomía, Veterinaria, Economía, Medicina, Psicología, etc. es muy común encontrar situaciones donde los datos recogidos son observaciones de **variables categóricas** cuyos niveles o categorías son empleados en la discriminación o identificación de las unidades muestrales en estudio. En este Capítulo se pretende introducir parcialmente el análisis de datos categóricos, el cual sólo se restringirá a la presentación del análisis de **tablas de contingencia.**

Definición 11.1: Variable categórica

Una variable categórica es una característica para la cual la escala de medida consiste de un conjunto de categorías.

En esta situación, los datos se presentan como frecuencias de observaciones que ocurren en la misma categoría.

Dentro de la escala categórica se distinguen tres tipos principales de variables:

Nominales: son aquellas cuyos niveles no están naturalmente ordenados, por ejemplo **color** del tegumento de semillas de maní, **variedad** de un cultivo, **raza** de animales, etc.

Ordinales: son aquellas cuyas distintas categorías tienen un orden natural, por ejemplo **grado** de ataque de una plaga (sin ataque, controlable, no controlable), **diagnóstico** de una enfermedad (seguro, probable, improbable), etc.

De intervalo: son aquellas variables de tipo numérico que tienen una distancia entre dos niveles, por ejemplo **edad** de los individuos (entre 15-20, 21-25 y 26-30 años), **diámetro** de los árboles (10-20, 21-30, 31-40 y 41-50 cm), etc.

Ordenando en forma decreciente los tipos de variables enunciados en función de la cantidad de información que proveen, se tiene: 1^o de intervalo, 2^o ordinal, 3^o nominal. Los métodos diseñados para un tipo de variable pueden ser usados para una de nivel superior. Así, una técnica para variables ordinales puede ser usada para una "de intervalo" pero no para una nominal. En este Capítulo no se enfatizarán los tipos diferentes de análisis ya que se necesitaría de una introducción a otros tópicos de modelación propiamente dichos, lo cual escapa a los objetivos de este libro.

Una variable puede ser nominal, ordinal o de intervalo, según lo que se mida o cómo se lo mida. Por ejemplo, la variable educación es nominal, si se refiere al tipo de educación: pública o privada; ordinal si mide el nivel de educación: preescolar, primario, secundario, terciario o universitario, mientras que es de intervalo si se cuantifica la cantidad de años de educación formal: 0, 1, 2,..., etc. (Agresti, 1990).

Cuando los individuos extraídos de una población son clasificados de acuerdo a, por lo menos, dos características observadas en ellos, se dice que los mismos están estudiándose en forma bivariada, esto es, por medio de dos variables aleatorias. Para analizar esa información se puede construir, entre otras cosas, una tabla de contingencia. Una tabla de contingencia se obtiene cuando el conjunto de individuos o entidades, como pueden ser semillas, personas, hojas, potreros, novillos, árboles, etc., son clasificados de acuerdo a uno o más criterios. Por ejemplo, las hojas de una hortaliza pueden ser clasificadas según tengan o no síntomas de enfermedad virósica y al mismo tiempo según provengan de la parte baja, media o alta de la planta.

Para el análisis de tablas de contingencia es necesario indagar primeramente en la clasificación de las variables que la definen. Ellas pueden ser: variables de *respuesta* o variables de *clasificación*. Las primeras, esto es las variables de respuesta o *dependientes*, son aleatorias y describen lo que fue observado en las unidades muestrales. Las segundas, las variables de clasificación o *independientes* o *factores*, son fijas por condicionamiento y las combinaciones de sus niveles definen estratos, poblaciones o subpoblaciones a las cuales las unidades muestrales pertenecen. De acuerdo con esta clasificación se definen dos tipos básicos de tablas de contingencia:

Tablas donde todas las variables son de *respuesta*;

Tablas donde algunas variables son de *respuesta* y otras de *clasificación*.

En el primer caso lo que interesa, usualmente, es verificar si existe asociación entre las variables, y cuando existe, construir algún coeficiente para medir ese grado de asociación. En el segundo caso, generalmente, el objetivo es estudiar los efectos de las variables de *clasificación* sobre la distribución conjunta de las variables de *respuesta* o sobre alguna característica específica de esa distribución. Un caso particular de gran importancia es aquel en que se considera sólo una variable de *respuesta* y las restantes como de *clasificación*. En este caso, como en el ANAVA, el objetivo es estudiar la influencia aislada o combinada de los factores en la distribución de la variable de respuesta.

Como ejemplos, obsérvese la Tabla 11.1 correspondiente a un ensayo cuyo objetivo era estudiar la influencia del estado del tegumento y la textura de la semilla en el éxito de la germinación de semillas de soja de la variedad Hood. O también los datos analizados por Grizzle *et al.* (1969), con la finalidad de indagar en la posible asociación entre severidad del ataque de una plaga y las prácticas culturales, como se muestra en Tabla 11.2. Considérense los datos de la Tabla 11.3, presentada por Birtlett (Díaz y De Luna, 1991), construida con el objetivo de estudiar el crecimiento de plantas sometidas a diferentes tratamientos. Las combinaciones de niveles de épocas de plantación con los niveles de las alturas de corte definen 4 subpoblaciones donde fue observada la distribución de la variable sobrevida. La Tabla 11.4, correspondiente a los resultados de un estudio sobre infecciones de querato-conjuntivitis en vacunos, donde el interés estuvo centrado en la verificación de interacción entre tratamientos y tipos de diagnóstico en relación a la proporción de curados. Se puede observar que el proceso de obtención de los datos combinados con los objetivos de la investigación permiten asociar a las observaciones 6 subpoblaciones (por condicionamiento) y dos categorías de respuesta, "curado" y "no curado".

Tabla 11.1: Clasificación de las observaciones realizadas en semillas de soja de la variedad Hood.

Tegumento	Germinó	Textura	Total
dañado	si	lisa	23
		rugosa	34
	no	lisa	109
		rugosa	78
sano	si	lisa	189
		rugosa	242
	no	lisa	56
		rugosa	69
total			800

Tabla 11.2: Número de plantas de maní según el grado de severidad de una plaga y práctica cultural del lote.

Prácticas Culturales	Severidad			Total
	Baja	Moderada	Alta	
Con rotación	235	124	38	397
Buena preparación de la cama de siembra	169	84	18	271
Uso de agroquímicos	452	67	27	546
Total	856	275	83	1214

Tabla 11.3: Número de plantas en función a la altura de corte y época de plantación.

Altura de corte	Época de plantación	Sobrevive	No Sobrevive	Total
Largo	Otras estaciones	156	84	240
	En primavera	84	156	240
Corto	Otras estaciones	107	133	240
	En primavera	31	209	240

Tabla 11.4: Número de animales según el tratamiento y el tipo de diagnóstico.

Tipo de Diagnóstico	Tratamiento	Curado	No curado
Complicado	A	78	20
	B	101	11
	C	68	46
Simple	A	40	5
	B	54	5
	C	34	6

Análisis de tablas de contingencia

Suponga que se lleva a cabo un estudio a campo, con plantas de soja, con el objetivo de evaluar el estado del cultivo en relación a la infestación de hongos y el tamaño de las plantas. La Tabla 11.5 contiene las observaciones recogidas en tal ensayo:

Tabla 11.5: Frecuencia de plantas en función de su sintomatología con respecto a la presencia de hongos.

Síntoma	Tamaño de la Planta			
	Alta	Media	Baja	Total
Enfermas	n_{11}	n_{12}	n_{13}	$n._{1}.$
Sanas	n_{21}	n_{22}	n_{23}	$n._{2}$
Total	$n._{1}$	$n._{2}$	$n._{3}$	$n..$

La presentación de la información en forma de tablas distingue básicamente la designación de filas (síntomas) y columnas (tamaño de la planta); el cuerpo de la tabla está constituido por:

a) celdas que contienen las frecuencias observadas n_{ij}

b) totales marginales de filas y columnas ($n_{i\bullet}$ y $n_{\bullet j}$ respectivamente), y total general ($n_{\bullet\bullet}$)

Una tabla que tiene R filas y C columnas se dice que es de dimensión R x C. En la Tabla 11.5 las variables categóricas, **X = síntomas** e **Y = tamaño**, con **dos** y **tres** niveles, respectivamente, conforman una tabla 2x3, esto es de **seis** celdas o combinaciones.

Si el análisis de tablas de contingencia tiene propósitos inferenciales es necesario considerar modelos probabilísticos para los datos. En este material no se profundizará en esta cuestión, simplemente se mencionan los principales modelos discretos, ya que la conformación de una tabla es a través de las frecuencias observadas. Entre los modelos más frecuentes se pueden mencionar: distribución Poisson (cuando el muestreo es aleatorio y no hay condicionamientos ni número total de observaciones fijado de antemano), distribución Multinomal (o Binomial para dos dimensiones), distribución Hipergeométrica, distribución Binomial Negativa y sus respectivos productos (cuando existe condicionamiento, esto es una de las variables es de clasificación y otra de respuesta).

En términos generales y de acuerdo con el tipo y número de variables involucradas en una tabla de contingencia se distinguen los tres casos principales que se presentan a continuación. Si bien las hipótesis que se desean probar en estos tres casos son diferentes el estadístico a usar es el mismo.

Tablas de contingencia a un criterio de clasificación

Si se extrae una muestra aleatoria simple de 100 semillas de un lote y se las clasifica según un criterio de calidad (como podría ser uno basado en la conductividad) en alta, media y baja, se obtiene una tabla con un único criterio de clasificación con tres niveles:

Calidad			
Alta	Media	Baja	Total
80	15	5	100

En el caso de que se dispusiera de alguna hipótesis sobre la distribución de la variable categórica *calidad de semilla*, estos resultados podrían utilizarse para someterla a prueba. Por ejemplo, si las especificaciones del lote de semillas, del cual se extrajo la muestra, dicen que las proporciones para las categorías de calidad son las siguientes:

Calidad		
Alta	Media	Baja
0.95	0.03	0.02

podría ser de interés probar si las frecuencias observadas son consistentes con las establecidas por las especificaciones, o no. Este tipo de análisis se conoce como **prueba de bondad de ajuste**.

Este enfoque también es utilizado en la siguiente situación. Algunas veces la variable en estudio es intrínsecamente continua o discreta, aunque por el método de observación seleccionado (o disponible) se la agrupa en clases convirtiéndola en una de intervalo. Ejemplos típicos ofrecen las variables como el diámetro de árboles, aumento de peso, etc. Si se desea verificar la hipótesis de que la muestra proviene de una distribución continua o discreta determinada (Normal, Poisson, etc.), las pruebas de bondad de ajuste se implementan de la misma manera que para las variables naturalmente categóricas.

La prueba de bondad de ajuste radica en la *comparación de las frecuencias* *observadas con aquellas esperadas (por un modelo)* mediante un *estadístico* conveniente. Cuando se realiza una prueba de bondad de ajuste, se establece como hipótesis nula que las frecuencias observadas (q_1 , q_2 ,...., q_k) son consistentes con las frecuencias esperadas (q_{10}, q_{20},...,q_{k0}). Para la construcción del estadístico se estiman las frecuencias esperadas cuando la Hipótesis Nula es cierta (q_{10}, q_{20},...,q_{k0}) y se calcula el estadístico:

$$\chi^2 = \frac{\sum_{i=1}^{k} (q_i - q_{i0})^2}{q_{i0}}$$

donde q_i es la i-ésima **frecuencia observada**, q_{i0} la i-ésima **frecuencia esperada** y **k** el **número de celdas** en la tabla. Bajo H_0 χ^2 se distribuye como una variable chi-cuadrado con ν grados de libertad.

Si χ^2 calculado es mayor que el cuantil (1-α) de la distribución chi-cuadrado con ν grados de libertad, entonces se rechaza H_0.

La clave para realizar las pruebas apropiadas en cada caso, reside en calcular correctamente las frecuencias esperadas bajo H_0 y los grados de libertad.

Considérese nuevamente el ejemplo de la calidad de semillas. Las frecuencias esperadas y observadas se muestran a continuación :

Frecuencias	Alta	Media	Baja	Total
Observadas	80	15	5	100
Esperadas	95	3	2	100

Los grados de libertad del estadístico χ^2 están dados por la diferencia entre el n*úmero de frecuencias esperadas necesarias para completar la tabla* **y** el *número de parámetros que deben estimarse para calcular dichas frecuencias (suponiendo cierta H_0).* En este ejemplo, la primera cantidad es 2 ya que, si el tamaño de la muestra está dado, conociendo las proporciones o frecuencias de 2 celdas la tercera es complementaria. Esto es, si en una muestra de tamaño 100 se conoce el número de semillas de alta y media calidad, el número de semillas de baja calidad queda inequívocamente establecido y de allí su proporción. La segunda cantidad, o sea el número de parámetros a estimar para completar la tabla cuando la hipótesis nula es cierta es, en este caso, 0. Esto se debe a que bajo H_0 todas las frecuencias están especificadas y no hace falta estimar ningún parámetro para calcularlas. Así, en este caso, los grados de libertad son 2-0=2.

Esta metodología analítica para la verificación de H_0 está sustentada en un teorema que establece la distribución asintótica del estadístico χ^2. El hecho de que la prueba se base en la distribución asintótica de un estadístico significa que tanto el nivel (α) como la potencia (1-β) serán aproximadamente los nominales (los que el investigador ha fijado o calculado) cuando n (el número total de observaciones) es grande (en el caso de igualdad de proporciones cada n_i debe ser grande). Por esta razón, la prueba χ^2 para tablas de contingencia debe ser usada con precaución cuando el/los tamaño/s de la/s muestra/s es/son pequeño/s. Así mismo, cuando más del 20% de las frecuencias esperadas son menores que 5, el estadístico χ^2 no ajusta a la distribución teórica y en consecuencia tanto el nivel como la potencia de la prueba se desconocen. En estos casos hay que recurrir a métodos exactos para calcular las probabilidades bajo H_0 como lo realiza la prueba de Irwin-Fisher para tablas de contingencia 2x2. Una discusión sobre estos métodos está fuera de los objetivos de esta presentación.

Tablas de contingencia a 2 criterios de clasificación (marginales libres)

Siguiendo el ejemplo anterior, una situación diferente se podría encontrar si se distinguieran, además, semillas claras y oscuras. Supóngase que las frecuencias obtenidas al clasificar 100 semillas, de un lote, por color y calidad fueron las siguientes:

Color	Calidad			
	Alta	Media	Baja	Total
Claras	16	3	1	20
Oscuras	64	12	4	80
Total	80	15	5	100

En este caso, como en el anterior, la tabla es en sí misma una herramienta descriptiva de la distribución de frecuencias y permite visualizar comportamientos que pueden ser de interés. Obsérvese que en la categoría semillas claras las frecuencias correspondientes a las calidades "media" y "baja" son relativamente menores (en términos absolutos) que las frecuencias correspondientes a dichas calidades en la categoría de semillas oscuras. El análisis correcto de esas tendencias implicará, obviamente, involucrar los marginales libres y por lo tanto otro modelo probabilístico para las observaciones.

Al igual que en el ejemplo anterior, el investigador podría contar con una distribución teórica para las categorías de calidad en cada una de las clases de color y el caso resultaría equivalente a un problema de bondad de ajuste (con 6 celdas). Sin embargo, lo usual, es que no se disponga de una distribución de frecuencias teórica en dos vías, por lo que se establece la **hipótesis de independencia**. Esta hipótesis establece que "la distribución de frecuencias para las calidades de semilla es la misma en ambas coloraciones" y viceversa; esto es, "la distribución de coloración es la misma independientemente de la calidad de semillas". Si la hipótesis de independencia no fuera cierta, entonces, se concluiría que la calidad de las semillas está asociada a la coloración. El análisis de esta hipótesis se conoce como **prueba Chi-cuadrado para la hipótesis de independencia,** ya que el estadístico de la prueba tiene distribución asintótica Chi-cuadrado.

Si la hipótesis nula se refiere a la independencia entre las dos variables de *respuesta* que conforman la tabla, esto implica que la distribución conjunta de las mismas puede obtenerse a partir del producto de las distribuciones marginales (es decir, la tendencia por fila - o por columna - es la misma en cada columna - o fila -). Pero, a diferencia del primer caso, las proporciones bajo H_0 no son conocidas y deben estimarse. Entonces, si H_0 es verdadera las proporciones para cada celda, por ejemplo la celda (ij)-ésima, está dada por el producto de las proporciones marginales. Para el ejemplo, la proporción esperada para semillas claras de alta calidad será $\frac{80}{100} \cdot \frac{20}{100}$ y la frecuencia esperada será entonces $\left(\frac{80}{100} \cdot \frac{20}{100}\right) \cdot 100$. Así se procede con las otras frecuencias esperadas para todas las celdas de la tabla.

En cuanto a los grados de libertad para el estadístico χ^2 ya definido, están dados por la diferencia entre el número de proporciones o frecuencias esperadas a especificar para completar la tabla y el número de parámetros a estimar bajo H_0. El primero de ellos, conocido el tamaño de la muestra (100), es 5, ya que una vez especificadas 5 proporciones, la restante queda establecida inequívocamente al construir la tabla. El segundo depende del número de filas y columnas de la tabla ya que, como se señaló en el cálculo de las frecuencias esperadas, sólo hacen falta conocer las frecuencias marginales. Así, si la tabla tiene 3 columnas, sólo hace falta conocer los totales de 2 de ellas, y si tiene 2 filas, sólo hace falta conocer el total de una de ellas. En total se necesitan conocer 3 frecuencias o proporciones (parámetros), por lo que los grados de libertad adecuados para la prueba son 5-3 = 2.

Tablas de Contingencia a 2 criterios de clasificación (marginales fijos)

Suponga que se muestrean, siguiendo el ejemplo anterior, 50 semillas claras y 50 semillas oscuras. Este esquema de muestreo difiere del caso anterior ya que ahora existe un factor de condicionamiento, la coloración. Antes se tomaba una muestra de 100 semillas sin tener en cuenta ninguna de sus características, generando una tabla con marginales libres. Ahora el muestreo para cada coloración de semilla genera una tabla con **marginales fijos** para las filas, como se muestra a continuación:

Color	Calidad				
	Alta	Media	Baja	Total	
Claras	15	25	10	50	*marginales*
Oscuras	35	10	5	50	*fijos*
Total	50	35	15	100	

marginales libres

Obsérvese que las filas resumen las distribuciones condicionales muestrales de la calidad de las semillas para cada coloración.

El interés es el mismo que en caso anterior, esto es establecer *si la calidad de las semillas está o no asociada a la coloración*. Reconociendo la generación de la tabla, es decir, cómo es recogida esa información, la hipótesis que se puede verificar es que *"las proporciones de cada clase de calidad son las mismas para cualquiera de las coloraciones"*. La prueba para contrastar esta hipótesis se conoce como **prueba Chi-cuadrado para la homogeneidad de proporciones**.

La hipótesis nula establece para este caso que las distribuciones condicionales de la variable utilizada como criterio columna respecto de aquella utilizada como criterio fila (en este caso, la variable con marginales fijos) son iguales. Esta hipótesis suele enunciarse como de igualdad de proporciones. Esto es, si se tiene una variable fila con k niveles y se toman muestras de tamaño $n_1, n_2, ..., n_k$ respectivamente, y si $(q_{110}, q_{120}, ..., q_{1p0})$, $(q_{210}, q_{220}, ..., q_{2p0})$,....., $(q_{k10}, q_{k20}, ..., q_{kp0})$ representan las proporciones verdaderas para los componentes columnas de cada una de las k filas, entonces estas k p-uplas son iguales bajo H_0.

Así, retomando el ejemplo anterior, como la hipótesis nula establece la igualdad de proporciones para la calidad de semillas, es necesario estimar las proporciones esperadas a partir de los datos como se procedió en el segundo caso. Recuérdese que antes el total de semillas era fijo siendo necesarias las frecuencias marginales de filas y columnas (distribuciones marginales) para estimar las frecuencias bajo H_0. En este tercer caso, los marginales fila son fijos y el estimador natural para la proporción de cada celda es la proporción basada en totales por columna (distribución incondicional) correspondiente. Luego, para la tabla presentada, la proporción esperada para semillas de alta calidad será $\frac{50}{100}$, y para semillas de baja calidad $\frac{15}{100}$ sin importar la coloración de las semillas. Por lo tanto sus frecuencias esperadas son $\frac{50}{100} \cdot 50$ y $\frac{15}{100} \cdot 50$ respectivamente. Así, la tabla de frecuencias esperadas será:

	Alta	Media	Baja	Total
Claras	25	17.5	7.5	50
Oscuras	25	17.5	7.5	50
Total	50	35	15	100

Para el cálculo de los grados de libertad, el número de proporciones a especificar para construir la tabla, dado que los marginales fila son fijos, es 4 (2 para cada coloración ya que la tercer celda de cada fila queda inequívocamente determinada). Por otro lado, el número de proporciones (parámetros) a especificar para construir la tabla bajo H_0 depende solamente de los marginales libres, y de los tres que presenta la tabla, sólo 2 son necesarios, por lo que los grados de libertad en esta tabla son 4 - 2 = 2.

Nota: Para las tablas a dos vías con r filas y c columnas (ya sea para el caso 2 o el caso 3), una regla práctica para calcular la frecuencia esperada para la celda-ij (fila i y columna j) es hacer el producto de los totales de la fila i y de la columna j ($n_{i.}$ y $n_{.j}$) y dividirlo por el total general n.. y para calcular los grados de libertad hacer el producto ν= (r-1) * (c-1).

A continuación se dan algunos ejemplos de aplicación:

Ejemplo 11.1(prueba de bondad de ajuste)

Un genetista realiza un cruzamiento de arvejas lisas y amarillas con arvejas rugosas y verdes, obteniendo los siguientes resultados:

Semillas	X_i
Lisas y amarillas	1080
Lisas y verdes	210
rugosas y amarillas	200
rugosas y verdes	110
Total	1600

Para saber si estas características siguen una de las leyes clásicas de la herencia *mendeliana* se trata de establecer si la frecuencia relativa de cada una de las clases en la población es: 9/16, 3/16, 3/16 y 1/16 respectivamente. Esta misma hipótesis se expresa como "la proporción es 9:3:3:1" (observar que 9 + 3 + 3 + 1 = 16, por lo que ambas formas son equivalentes). Así

H_0: la frecuencia es 9:3:3:1 versus H_1: la frecuencia no es 9:3:3:1

Los valores esperados, si la hipótesis nula es cierta, surgen de multiplicar cada una de las frecuencias relativas (o proporciones) por el total de individuos observados en la muestra. Por lo tanto la tabla de frecuencias esperadas es:

Tipo	Esperadas
lisas y amarillas	9/16*1600 = 900
lisas y verdes	3/16*1600 = 300
rugosas y amarillas	3/16*1600 = 300
rugosas y verdes	1/16*1600 = 100
Total	1600

El estadístico descripto como χ^2 tiene una distribución aproximada χ^2con (r-1-k) grados de libertad, (r es la cantidad de categorías y k la cantidad de parámetros estimados). En el ejemplo, r = 4, k = 0, luego $\chi^2 \sim \chi^2_3$. El estadístico evaluado en este caso es:

$$\chi^2 = \frac{(1080 - 900)^2}{900} + \frac{(210\text{-}300)^2}{300} + \frac{(200\text{-}300)^2}{300} + \frac{(110 - 100)^2}{100} = 97.33$$

La región de rechazo para este contraste está siempre a la derecha, o sea, para valores grandes de χ^2. El cuantil (1-α) de la distribución χ^2 con 3 grados de libertad es 7.81 para $\alpha = 0.05$. Como 97.3 es mayor que 7.81, se rechaza H_0; las frecuencias no siguen una distribución 9:3:3:1.

Ejemplo 11.2: (prueba de independencia)

Una forma intuitiva de considerar la falta de independencia entre dos variables es pensar que si se conoce la modalidad de una de ellas, entonces se conoce la probabilidad de ocurrencia de distintas modalidades de la otra variable en la misma observación. En este caso no hay independencia o sea, ambas variables están estadísticamente asociadas.

Una muestra aleatoria de 1260 semillas fue extraída para estudiar su textura (lisa, intermedia, rugosa) y su velocidad de germinación (alta, media, baja, nula). Se construye con esa información la siguiente tabla de contingencia:

	Textura			
Germinación	Lisa	Intermedia	Rugosa	Total
Alta	122	30	20	172
Media	226	51	66	343
Baja	306	115	96	517
Nula	131	59	38	228
Total	785	225	220	1260

la hipótesis nula que interesa probar es si la variable textura es independiente de la variable velocidad de germinación, por lo tanto se tiene:

H_0: Hay independencia entre las variables germinación y textura, versus

H_1: No hay independencia entre las variables germinación y textura

Si la hipótesis nula es cierta, las frecuencias esperadas se calculan según se describió anteriormente y la tabla es:

	Textura			
Germinación	Lisa	Intermedia	Rugosa	Total
Alta	107.16	34.81	30.03	172
Media	213.69	69.42	59.88	343
Baja	322.10	104.63	90.27	517
Nula	142.05	46.14	39.81	228
Total	785	225	220	1260

luego, evaluando el estadístico se tiene:

$$\chi^2 = \frac{(122 - 107.16)^2}{107.16} + \ldots + \frac{(38 - 39.81)^2}{39.81} = 18.24$$

Como $\chi^2 = 18.24$ es mayor que 12.6, que es el cuantil 1 - α de la distribución χ^2 con 6 ([4-1]x[3-1]) grados de libertad y $\alpha = 0.05$, se rechaza H_0, o sea, no hay independencia entre las variables.

Ejemplo 11.3: (homogeneidad de proporciones)

Se desea conocer si la proporción de pulgones muertos después de ser tratados con distintas dosis de un insecticida es la misma o no. Para ello se toman 3 muestras aleatorias de 100 pulgones cada una y se las asigna al azar a los tratamientos consistentes en aplicaciones del insecticida en dosis de 20 ppm., 40 ppm. y 80 ppm. Los resultados fueron:

	Dosis			
Condición	20 ppm.	40 ppm.	80 ppm.	Total
Muertos	32	59	92	183
Vivos	68	41	8	117
Total	100	100	100	300

La hipótesis nula expresa que las proporciones de insectos muertos con 20, 40 y 80 ppm. son iguales. Si esta hipótesis es cierta, se debería esperar que la cantidad de insectos muertos en los tres tratamientos sea proporcional a la cantidad de individuos en cada tratamiento (marginales fijos de columnas). La tabla de frecuencias esperadas es:

	Dosis			
Condición	20 ppm	40 ppm	80 ppm	Total
Muertos	61	61	61	183
Vivos	39	39	39	117
Total	100	100	100	300

Cuando la hipótesis nula es verdadera y cada n_{ij} es suficientemente grande, el estadístico χ^2 tiene una distribución aproximada χ^2 con (r-1)(c-1) grados de libertad. Así, el estadístico evaluado es:

$$\chi^2 = \frac{(32 - 61)^2}{61} + \frac{(59 - 61)^2}{61} + \frac{(92 - 61)^2}{61} + \frac{(68 - 39)^2}{39} + \frac{(41 - 39)^2}{39} + \frac{(8 - 39)^2}{39} = 75.914$$

Como $\chi^2 \sim \chi^2_2$, la región de rechazo para $\alpha = 0.05$ está dada por aquellos valores de $\chi^2 > \chi^2_{2;\,0.95}$ es decir, los valores de $\chi^2 > 5.99$. En este caso $\chi^2 > 5.99$, y por lo tanto se rechaza H_0, por lo que no todas las dosis tienen el mismo efecto sobre la mortalidad de los pulgones. En otras palabras lo que este test dice es que lo observado no es atribuido al azar sino al efecto de un tratamiento (Dosis) y de allí el sentido de su aplicación.

Ejercicios

Ejercicio 11.1

Un estudio diagnóstico fue llevado a cabo a los fines de indagar sobre la existencia de asociación entre el tipo de pérdidas de un cultivo y dos métodos de aplicación de un fungicida. Los resultados siguientes resumen la información de 22071 lotes de cultivos en la región pampeana del país.

	Tipo de Pérdida		
Método	total	moderada	sin pérdidas
Tradicional	18	171	10845
No tradicional	5	99	10933

a) ¿Cuál es la hipótesis estadística a evaluar?

b) Realizar el análisis para la verificación de dicha hipótesis y concluir.

Ejercicio 11.2

Se observaron 80 nacimientos obtenidos del cruzamiento de 10 chanchas con el mismo padrillo, de los cuales 42 fueron rojizos, 12 negros y 26 blancos. El modelo genético supuesto en este cruzamiento prevé una distribución de colores con frecuencias 9:3:4.

¿Son los datos consistentes con el modelo teórico propuesto al nivel de significación del 0.01?

Ejercicio 11.3

Una fábrica de implementos agrícolas desea determinar si las causas de ausentismo se relacionan con la edad. Se tomó una muestra de 200 empleados al azar y se clasificaron según edad y causa de ausentismo:

Edad	Menos de 30	30 a 50	Más de 50
Enfermedad	40	28	52
Otras	20	36	24

¿Qué contraste se puede realizar? Trabajar con un $\alpha = 0.01$

Ejercicio 11.4

Se dispone de 300 animales de laboratorio y se decide tratar a 200 con una vacuna experimental y dejar a 100 como controles. Después de tratar al primer lote se expone a los 300 al contagio de la enfermedad en estudio. El recuento final, después de un período experimental adecuado, fue:

	Enfermos	Sanos	Total
Tratados	56	144	200
No Tratados	71	29	100
Total	127	173	300

¿Qué tipo de contraste se puede realizar?

Ejercicio 11.5

En un cruzamiento de híbridos de tomate se obtuvieron los siguientes valores:

Fenotipos	Frecuencias Observadas
Alto, sin brotes	926
Alto, con brotes	288
Bajo, sin brotes	293
Bajo, con brotes	104

¿Corroboran estos datos la proporción 9:3:3:1, con $\alpha = 0.05$?

Ejercicio 11.6

Un ecólogo deseaba estudiar si había relación entre la textura de la hoja de una especie rara y el tipo de suelo donde crecía. Para ello tomó una superficie de 400 km^2, y seleccionó una muestra al azar de 100 árboles obteniendo los siguientes resultados:

	Textura	
Tipo de Suelo	Con Pelusa	Lisa
Suelo calcáreo	12	22
Suelo no calcáreo	16	50

a) ¿Qué hipótesis estadística se puede probar?

b) Realizar la prueba correspondiente con nivel de significación $\alpha = 0.05$

c) ¿Qué conclusiones se pueden extraer?

Bibliografía

1. AGRESTI, A (1990). Categorical Data Analysis. *John Wiley & Sons.* 558 p.
2. BAHRD, J. C. (1970). Psichophysical Analysis of Visual Space. *Pergamon Press, New York.*
3. BERENSON, M.L.; LEVINE, D.M. and M. GOLDSTEIN. (1983). Intermediate Statistical Methods and Applications. A Computer Package Approach. *Prentice-Hall, Inc., Englewood Cliffs, New Jersey*. 568 p.
4. BIANCO, A. (1994). Introducción al Análisis de Datos Categóricos. *Coloquio Unión Matemática Argentina (UMA).* La Falda 55pp.
5. CASANOVES F., DI RIENZO J.A., ROBLEDO C.W. (1998). Bases para estadística experimental. Editorial Screen, Córdoba, Agentina. 189 p.
6. CLEVELAND, W. S. (1984). Graphs in scientific publications*. The American Statistician*, 38: 261-269.
7. CLEVELAND, W. S. (1985). The elements of graphing data. *Wadsworth Advanced Books. Monterrey, CA*.
8. COHRAN, W.G. (1981). Técnicas de Muestreo*. Compañía Editorial Continental S. A. México*. 513 p.
9. CONOVER, W.J. (1980). Practical Nonparametric Statistics 2^{ed} . John Wiley & Sons. New York. Pp 485.
10. CHING CHUN LI. (1977). Introducción a la Estadística Experimental. *Ediciones Omega, S.A. Barcelona*. 496 p.
11. DIAZ, M.P. y De LUNA, J.G. (1991). Análise de dados categorizados em tabelas de contingencia LxC. Seminario Curso de Post-Grado. Escuela Superior de Agricultura Luis de Queiroz. Universidad de Sao Paulo. Piracicaba, SP. Brasil. 87p.
12. DIXON, W. and F.J. MASSEY. (1970). Introducción al Análisis Estadístico. *McGraw-Hill, Inc. México*. 489 p.
13. DRAPER N.R, SMITH H. (1998). Applied Regression Analysis (Third Edition). John Wiley & Sons, New York 706 p.
14. DUNCAN, D.B. (1955). Multiple and multiple F-test. *Biometrics,* 11:1-42.

15. DUNNETT, C.W. (1964). New tables for multiple comparisons with the control. *Biometrics*, 20:482-491.
16. FISHER, R.A. (1966). The design of experiments. 8th Edition. *Hafner Publishing Co. New York*.
17. FREUND, J.E. and R.E. WALPOLE. (1980). Mathematical Statistics. Third edition. *Prentice-Hall, Inc., Englewood Cliffs, New Jersey*. 450 p.
18. GILBERT, N. (1980). Estadística. *Nueva Editorial Interamericana. S. A. de C. V.* 346 p.
19. GRIZZLE, J.E., C.F. STARMER, and G.G. KOCH. (1969). Analisys of categorical data by linear models. *Biometrics*, 25: 489-504.
20. HINKLEY, D. V. (1992). Cox, Dr. Theoritical Statistics, *Chapema of Hall, London*. 511 p.
21. JAMES, B. (1981). Probabilidade: Un curso en nivel intermediario. *CNPq, Brasilia* 293 p.
22. KOLMOGOROFF, A. N. (1937). Citeral in Kolmogoroff, A. N. and Fowin, S. V. (1961). Measure, Lekesgure Integrals, and Hilbert Space. *Academic Press, New York*.
23. KOSSLYN, S. M. (1980). Image and Mind. *Cambridge, M.A.: Hardvard University Press*.
24. MILLIKEN G.A. and JOHNSON, D.E. (1995) Analysis of Messy Data. Volume I: Designed Experiments. *Van Nostrand Reinhold, NewYork*.
25. MONTGOMERY, D.C. (1991). Diseño y Análisis de Experimentos. *Grupo Editorial Iberoamérica.* 589 p.
26. MOOD, A.; GRAYBILL, F.; BOES, D. C. (1974) Introduction to the theory of Statistics. *Mc Graw-Hill, Inc*. 3th. 564 p.
27. MOSTELLER, F. and J. W. TUKEY. (1977). Data Analysis and Regression. *Addison-Westley Publishing Company*. 588 p.
28. MYERS, R.H. (1990). Classical and modern Regression with Applications. *P. W. S.-Kent Publishing Company*. 488 p.
29. PIMENTEL GOMES, F. (1978). Curso de Estadística Experimental. *Editorial Hemisferio Sur S.A.* 323 p.
30. RAWLINGS, J.O. (1988). Applied Regression Analisys: a Research Tool.

Wadsworth & Brooks/ Cole Advance Books & Software. 552 p.

31. SCHEFFE, H. (1953). A method for judging all contrasts in the analysis of variance. *Biometrika,* 40:87-104.
32. SHEPARD, R. N. (1978). The mental image. American *Psychologist*, 33: 125-127.
33. SOKAL, R. and J. ROHLF. (1981). Biometry: the Principles and Practices of statistics in Biological Research. *W. H. Sreeman and Company*. 587 p.
34. STEEL, R.G.D. y J.H. TORRIE. (1985). Bioestadística: Principios y Procedimientos. *Mc Graw-Hill*. 622 p.
35. SYSTAT. (1977). The System for Statistics. *Systat Inc*. 415 p.
36. TUKEY, J.W. (1949). Comparing individual means in the analysis of variance. *Biometrics*, 5:99-114.

Tablas Estadísticas

Tabla de Números Aleatorios

81	4	37	23	59	51	32	71	89	37	66	28	38	49	59	49	33	77	42
82	24	34	34	71	62	74	66	32	26	75	20	47	68	86	92	81	19	9
73	34	62	51	22	38	24	28	45	44	25	68	74	68	26	64	44	79	94
76	27	21	30	62	52	44	30	84	6	44	60	31	31	39	4	18	33	59
4	48	54	8	86	7	43	52	86	63	84	74	72	91	29	96	73	5	60
18	10	75	64	40	44	2	66	24	45	58	44	73	79	66	95	25	49	80
34	100	36	14	79	51	49	35	93	97	28	4	78	2	34	58	40	9	48
53	46	39	11	61	33	12	8	70	28	2	7	87	58	7	59	2	68	48
79	25	52	36	53	64	29	57	84	26	56	11	15	69	52	42	20	12	99
66	10	24	92	19	74	100	85	39	5	39	39	58	8	49	34	41	77	70
99	84	99	91	41	88	9	33	24	99	96	98	18	89	44	93	12	17	92
50	28	33	52	84	40	21	5	49	92	21	31	2	62	53	13	96	69	85
76	55	77	53	13	39	64	43	58	64	31	78	56	95	49	57	2	64	56
93	35	75	28	48	100	98	48	27	12	94	27	84	43	32	18	19	13	77
7	17	21	49	100	15	59	83	10	67	99	4	26	88	33	27	80	63	73
72	38	80	72	69	22	19	17	65	68	66	84	83	97	86	8	55	74	93
7	5	58	68	42	70	2	16	23	35	60	45	35	60	43	62	69	7	58
19	34	58	54	20	91	95	72	16	37	46	57	93	31	97	2	96	81	6
40	72	65	99	49	40	10	68	88	14	11	84	22	91	55	44	79	85	84
99	37	83	34	31	43	86	58	30	67	21	2	54	27	46	11	32	43	10
2	16	91	60	88	6	26	5	58	44	97	90	90	28	12	78	67	45	5
80	7	47	41	67	64	96	49	84	42	87	33	15	28	58	64	42	49	74
53	20	35	44	18	26	47	6	1	55	6	74	62	56	23	51	78	15	19
73	88	60	42	74	2	31	32	85	40	21	42	68	35	51	58	87	5	10
32	13	59	78	14	50	89	18	41	63	35	49	67	72	31	66	79	22	14
67	51	56	9	52	98	83	41	16	43	50	27	94	48	66	6	20	43	23
95	52	3	87	98	43	17	72	50	58	31	27	92	46	31	69	72	67	27
45	67	22	41	55	27	32	44	80	34	57	10	37	30	5	65	59	27	99
82	63	70	7	59	37	61	58	99	31	33	69	10	79	32	50	56	48	78
97	50	13	19	83	27	23	55	88	57	67	8	58	76	56	62	15	76	56
46	37	31	68	62	89	98	57	60	70	24	76	44	57	86	62	83	26	59
76	22	34	79	33	45	32	43	76	7	45	12	61	24	29	20	24	45	65
44	94	14	84	72	5	19	19	61	47	18	21	41	96	17	45	63	5	6
20	65	87	43	77	46	73	38	74	18	73	62	25	18	24	68	27	64	51
34	14	3	89	68	56	33	33	67	14	9	38	58	95	32	14	54	34	65
13	80	93	61	53	61	95	63	35	52	80	83	84	61	25	76	20	13	73
35	98	76	30	2	7	1	88	19	9	39	44	39	38	40	42	60	15	10
81	33	39	20	88	46	73	62	41	93	49	53	48	40	17	40	83	12	53
19	26	69	65	72	64	9	28	14	75	57	35	25	90	49	23	83	71	30
63	36	77	14	9	94	59	3	16	100	89	93	93	97	4	69	90	97	40
53	44	47	62	82	41	77	18	59	65	31	86	41	39	78	77	24	65	79
15	63	14	64	93	89	55	27	46	27	67	38	38	26	94	24	82	86	63
85	13	32	99	4	4	46	40	95	10	33	30	98	3	53	17	86	63	93
5	83	68	8	51	95	7	37	42	38	57	99	58	74	53	42	67	1	68
49	19	61	29	69	26	39	58	4	42	22	11	99	2	53	17	13	76	5
83	76	63	26	32	66	42	55	85	15	72	78	27	51	25	82	71	38	13
58	24	35	54	45	36	69	36	41	92	85	16	59	99	99	12	58	19	51
29	45	5	17	94	51	56	13	55	79	39	18	62	58	9	59	36	46	45
87	4	54	61	45	75	31	68	92	96	51	76	20	41	28	80	69	88	84
95	4	25	62	86	89	90	88	21	66	33	32	6	59	82	3	67	41	44
4	44	99	80	20	29	89	21	44	33	85	77	25	26	40	50	25	47	77
34	78	11	64	83	68	5	56	53	34	32	14	90	31	57	47	82	84	31
33	23	22	97	13	28	2	91	85	67	49	41	81	74	94	28	49	82	25
56	14	92	52	25	15	60	46	29	5	54	91	58	19	88	15	29	86	36
43	77	74	77	84	66	49	38	72	84	86	77	9	4	26	69	38	65	31

Tabla de Cuantiles de la una Distribución Normal Estándar

z	P(Z ≤ z)	z	P(Z ≤ z)	z	P(Z ≤ z)	quantil	z
-3.25	0.00058	-1.00	0.15866	1.25	0.89435	0.00001	-4.265
-3.20	0.00069	-0.95	0.17106	1.30	0.90320	0.0001	-3.719
-3.15	0.00082	-0.90	0.18406	1.35	0.91149	0.001	-3.090
-3.10	0.00097	-0.85	0.19766	1.40	0.91924	0.005	-2.576
-3.05	0.00114	-0.80	0.21186	1.45	0.92647	0.01	-2.326
-3.00	0.00135	-0.75	0.22663	1.50	0.93319	0.02	-2.054
-2.95	0.00159	-0.70	0.24196	1.55	0.93943	0.025	-1.960
-2.90	0.00187	-0.65	0.25785	1.60	0.94520	0.03	-1.881
-2.85	0.00219	-0.60	0.27425	1.65	0.95053	0.04	-1.751
-2.80	0.00256	-0.55	0.29116	1.70	0.95543	0.05	-1.645
-2.75	0.00298	-0.50	0.30854	1.75	0.95994	0.06	-1.555
-2.70	0.00347	-0.45	0.32636	1.80	0.96407	0.07	-1.476
-2.65	0.00402	-0.40	0.34458	1.85	0.96784	0.08	-1.405
-2.60	0.00466	-0.35	0.36317	1.90	0.97128	0.09	-1.341
-2.55	0.00539	-0.30	0.38209	1.95	0.97441	0.10	-1.282
-2.50	0.00621	-0.25	0.40129	2.00	0.97725	0.15	-1.036
-2.45	0.00714	-0.20	0.42074	2.05	0.97982	0.20	-0.842
-2.40	0.00820	-0.15	0.44038	2.10	0.98214	0.25	-0.674
-2.35	0.00939	-0.10	0.46017	2.15	0.98422	0.30	-0.524
-2.30	0.01072	-0.05	0.48006	2.20	0.98610	0.35	-0.385
-2.25	0.01222	0.00	0.50000	2.25	0.98778	0.40	-0.253
-2.20	0.01390	0.05	0.51994	2.30	0.98928	0.45	-0.126
-2.15	0.01578	0.10	0.53983	2.35	0.99061	0.50	0.000
-2.10	0.01786	0.15	0.55962	2.40	0.99180	0.55	0.126
-2.05	0.02018	0.20	0.57926	2.45	0.99286	0.60	0.253
-2.00	0.02275	0.25	0.59871	2.50	0.99379	0.65	0.385
-1.95	0.02559	0.30	0.61791	2.55	0.99461	0.70	0.524
-1.90	0.02872	0.35	0.63683	2.60	0.99534	0.75	0.674
-1.85	0.03216	0.40	0.65542	2.65	0.99598	0.80	0.842
-1.80	0.03593	0.45	0.67364	2.70	0.99653	0.85	1.036
-1.75	0.04006	0.50	0.69146	2.75	0.99702	0.90	1.282
-1.70	0.04457	0.55	0.70884	2.80	0.99744	0.91	1.341
-1.65	0.04947	0.60	0.72575	2.85	0.99781	0.92	1.405
-1.60	0.05480	0.65	0.74215	2.90	0.99813	0.93	1.476
-1.55	0.06057	0.70	0.75804	2.95	0.99841	0.94	1.555
-1.50	0.06681	0.75	0.77337	3.00	0.99865	0.95	1.645
-1.45	0.07353	0.80	0.78814	3.05	0.99886	0.96	1.751
-1.40	0.08076	0.85	0.80234	3.10	0.99903	0.97	1.881
-1.35	0.08851	0.90	0.81594	3.15	0.99918	0.975	1.960
-1.30	0.09680	0.95	0.82894	3.20	0.99931	0.98	2.054
-1.25	0.10565	1.00	0.84134	3.25	0.99942	0.99	2.326
-1.20	0.11507	1.05	0.85314	3.30	0.99952	0.995	2.576
-1.15	0.12507	1.10	0.86433	3.35	0.99960	0.999	3.090
-1.10	0.13567	1.15	0.87493	3.40	0.99966	0.9999	3.719
-1.05	0.14686	1.20	0.88493	3.45	0.99972	0.99999	4.265

Tabla de Cuantiles de la Distribución T de Student

En el margen superior se leen los cuatiles y en el margen izquierdo los grados de libertad (ν). Esta tabla tabula valores $P(T \leq t)$ para $t>0$. Si se buscan valores de $t<0$ los cuantiles se leen en el margen inferior.

ν	0.700	0.725	0.750	0.775	0.800	0.825	0.850	0.875	0.900	0.925	0.950	0.975	0.990	0.995
1	0.727	0.854	1.000	1.171	1.376	1.632	1.963	2.414	3.078	4.165	6.314	12.71	31.82	63.66
2	0.617	0.713	0.816	0.931	1.061	1.210	1.386	1.604	1.886	2.282	2.920	4.303	6.965	9.925
3	0.584	0.671	0.765	0.866	0.978	1.105	1.250	1.423	1.638	1.924	2.353	3.182	4.541	5.841
4	0.569	0.652	0.741	0.836	0.941	1.057	1.190	1.344	1.533	1.778	2.132	2.776	3.747	4.604
5	0.559	0.641	0.727	0.819	0.920	1.031	1.156	1.301	1.476	1.699	2.015	2.571	3.365	4.032
6	0.553	0.633	0.718	0.808	0.906	1.013	1.134	1.273	1.440	1.650	1.943	2.447	3.143	3.707
7	0.549	0.628	0.711	0.800	0.896	1.001	1.119	1.254	1.415	1.617	1.895	2.365	2.998	3.499
8	0.546	0.624	0.706	0.794	0.889	0.993	1.108	1.240	1.397	1.592	1.860	2.306	2.896	3.355
9	0.543	0.621	0.703	0.790	0.883	0.986	1.100	1.230	1.383	1.574	1.833	2.262	2.821	3.250
10	0.542	0.619	0.700	0.786	0.879	0.980	1.093	1.221	1.372	1.559	1.812	2.228	2.764	3.169
11	0.540	0.617	0.697	0.783	0.876	0.976	1.088	1.214	1.363	1.548	1.796	2.201	2.718	3.106
12	0.539	0.615	0.695	0.781	0.873	0.972	1.083	1.209	1.356	1.538	1.782	2.179	2.681	3.055
13	0.538	0.614	0.694	0.779	0.870	0.969	1.079	1.204	1.350	1.530	1.771	2.160	2.650	3.012
14	0.537	0.613	0.692	0.777	0.868	0.967	1.076	1.200	1.345	1.523	1.761	2.145	2.624	2.977
15	0.536	0.612	0.691	0.776	0.866	0.965	1.074	1.197	1.341	1.517	1.753	2.131	2.602	2.947
16	0.535	0.611	0.690	0.774	0.865	0.963	1.071	1.194	1.337	1.512	1.746	2.120	2.583	2.921
17	0.534	0.610	0.689	0.773	0.863	0.961	1.069	1.191	1.333	1.508	1.740	2.110	2.567	2.898
18	0.534	0.609	0.688	0.772	0.862	0.960	1.067	1.189	1.330	1.504	1.734	2.101	2.552	2.878
19	0.533	0.609	0.688	0.771	0.861	0.958	1.066	1.187	1.328	1.500	1.729	2.093	2.539	2.861
20	0.533	0.608	0.687	0.771	0.860	0.957	1.064	1.185	1.325	1.497	1.725	2.086	2.528	2.845
21	0.532	0.608	0.686	0.770	0.859	0.956	1.063	1.183	1.323	1.494	1.721	2.080	2.518	2.831
22	0.532	0.607	0.686	0.769	0.858	0.955	1.061	1.182	1.321	1.492	1.717	2.074	2.508	2.819
23	0.532	0.607	0.685	0.769	0.858	0.954	1.060	1.180	1.319	1.489	1.714	2.069	2.500	2.807
24	0.531	0.606	0.685	0.768	0.857	0.953	1.059	1.179	1.318	1.487	1.711	2.064	2.492	2.797
25	0.531	0.606	0.684	0.767	0.856	0.952	1.058	1.178	1.316	1.485	1.708	2.060	2.485	2.787
26	0.531	0.606	0.684	0.767	0.856	0.952	1.058	1.177	1.315	1.483	1.706	2.056	2.479	2.779
27	0.531	0.605	0.684	0.767	0.855	0.951	1.057	1.176	1.314	1.482	1.703	2.052	2.473	2.771
28	0.530	0.605	0.683	0.766	0.855	0.950	1.056	1.175	1.313	1.480	1.701	2.048	2.467	2.763
29	0.530	0.605	0.683	0.766	0.854	0.950	1.055	1.174	1.311	1.479	1.699	2.045	2.462	2.756
30	0.530	0.605	0.683	0.765	0.854	0.949	1.055	1.173	1.310	1.477	1.697	2.042	2.457	2.750
31	0.530	0.604	0.682	0.765	0.853	0.949	1.054	1.172	1.309	1.476	1.696	2.040	2.453	2.744
32	0.530	0.604	0.682	0.765	0.853	0.948	1.054	1.172	1.309	1.475	1.694	2.037	2.449	2.738
33	0.530	0.604	0.682	0.765	0.853	0.948	1.053	1.171	1.308	1.474	1.692	2.035	2.445	2.733
34	0.529	0.604	0.682	0.764	0.852	0.948	1.052	1.170	1.307	1.473	1.691	2.032	2.441	2.728
35	0.529	0.604	0.682	0.764	0.852	0.947	1.052	1.170	1.306	1.472	1.690	2.030	2.438	2.724
36	0.529	0.603	0.681	0.764	0.852	0.947	1.052	1.169	1.306	1.471	1.688	2.028	2.434	2.719
37	0.529	0.603	0.681	0.764	0.851	0.947	1.051	1.169	1.305	1.470	1.687	2.026	2.431	2.715
38	0.529	0.603	0.681	0.763	0.851	0.946	1.051	1.168	1.304	1.469	1.686	2.024	2.429	2.712
39	0.529	0.603	0.681	0.763	0.851	0.946	1.050	1.168	1.304	1.468	1.685	2.023	2.426	2.708
40	0.529	0.603	0.681	0.763	0.851	0.946	1.050	1.167	1.303	1.468	1.684	2.021	2.423	2.704
41	0.529	0.603	0.681	0.763	0.850	0.945	1.050	1.167	1.303	1.467	1.683	2.020	2.421	2.701
42	0.528	0.603	0.680	0.763	0.850	0.945	1.049	1.166	1.302	1.466	1.682	2.018	2.418	2.698
43	0.528	0.603	0.680	0.762	0.850	0.945	1.049	1.166	1.302	1.466	1.681	2.017	2.416	2.695
44	0.528	0.602	0.680	0.762	0.850	0.945	1.049	1.166	1.301	1.465	1.680	2.015	2.414	2.692
45	0.528	0.602	0.680	0.762	0.850	0.944	1.049	1.165	1.301	1.465	1.679	2.014	2.412	2.690
46	0.528	0.602	0.680	0.762	0.850	0.944	1.048	1.165	1.300	1.464	1.679	2.013	2.410	2.687
47	0.528	0.602	0.680	0.762	0.849	0.944	1.048	1.165	1.300	1.463	1.678	2.012	2.408	2.685
48	0.528	0.602	0.680	0.762	0.849	0.944	1.048	1.164	1.299	1.463	1.677	2.011	2.407	2.682
49	0.528	0.602	0.680	0.762	0.849	0.944	1.048	1.164	1.299	1.462	1.677	2.010	2.405	2.680
50	0.528	0.602	0.679	0.761	0.849	0.943	1.047	1.164	1.299	1.462	1.676	2.009	2.403	2.678
	0.300	0.275	0.250	0.225	0.200	0.175	0.150	0.125	0.100	0.075	0.050	0.025	0.010	0.005

Tabla de Cuantiles de la Distribución Chi-Cuadrado

En el margen superior se lee $P(\chi^2 \leq x)$ para los valores de x que figuran en el cuerpo de la tabla y en el margen izquierdo los grados de libertad (ν).

ν	0.010	0.025	0.05	0.10	0.15	0.20	0.25	0.30	0.35	0.40	0.45	0.50
1	0.0002	0.0010	0.0039	0.0158	0.0358	0.0642	0.1015	0.1485	0.2059	0.2750	0.3573	0.4549
2	0.0201	0.0506	0.1026	0.2107	0.3250	0.4463	0.5754	0.7133	0.8616	1.0217	1.1957	1.3863
3	0.1148	0.2158	0.3518	0.5844	0.7978	1.0052	1.2125	1.4237	1.6416	1.8692	2.1095	2.3660
4	0.2971	0.4844	0.7107	1.0636	1.3665	1.6488	1.9226	2.1947	2.4701	2.7528	3.0469	3.3567
5	0.5543	0.8312	1.1455	1.6103	1.9938	2.3425	2.6746	2.9999	3.3251	3.6555	3.9959	4.3515
6	0.8721	1.2373	1.6354	2.2041	2.6613	3.0701	3.4546	3.8276	4.1973	4.5702	4.9519	5.3481
7	1.2390	1.6899	2.1674	2.8331	3.3583	3.8223	4.2549	4.6713	5.0816	5.4932	5.9125	6.3458
8	1.6465	2.1797	2.7326	3.4895	4.0782	4.5936	5.0706	5.5274	5.9753	6.4226	6.8766	7.3441
9	2.0879	2.7004	3.3251	4.1682	4.8165	5.3801	5.8988	6.3933	6.8763	7.3570	7.8434	8.3428
10	2.5582	3.2470	3.9403	4.8652	5.5701	6.1791	6.7372	7.2672	7.7832	8.2955	8.8123	9.3418
11	3.0535	3.8157	4.5748	5.5778	6.3364	6.9887	7.5841	8.1479	8.6952	9.2373	9.7831	10.3410
12	3.5706	4.4038	5.2260	6.3038	7.1138	7.8073	8.4384	9.0343	9.6115	10.1820	10.7553	11.3403
13	4.1069	5.0088	5.8919	7.0415	7.9008	8.6339	9.2991	9.9257	10.5315	11.1291	11.7288	12.3398
14	4.6604	5.6287	6.5706	7.7895	8.6963	9.4673	10.1653	10.8215	11.4548	12.0785	12.7034	13.3393
15	5.2294	6.2621	7.2610	8.5468	9.4993	10.3070	11.0365	11.7212	12.3809	13.0297	13.6790	14.3389
16	5.8122	6.9076	7.9616	9.3122	10.3090	11.1521	11.9122	12.6244	13.3096	13.9827	14.6555	15.3385
17	6.4078	7.5642	8.6718	10.0852	11.1249	12.0023	12.7919	13.5307	14.2406	14.9373	15.6328	16.3382
18	7.0149	8.2307	9.3905	10.8649	11.9462	12.8570	13.6753	14.4399	15.1738	15.8932	16.6108	17.3379
19	7.6327	8.9065	10.1170	11.6509	12.7727	13.7158	14.5620	15.3517	16.1089	16.8504	17.5894	18.3377
20	8.2604	9.5908	10.8508	12.4426	13.6039	14.5784	15.4518	16.2659	17.0458	17.8088	18.5687	19.3374
21	8.8972	10.2829	11.5913	13.2396	14.4393	15.4446	16.3444	17.1823	17.9843	18.7683	19.5485	20.3372
22	9.5425	10.9823	12.3380	14.0415	15.2788	16.3140	17.2396	18.1007	18.9243	19.7288	20.5288	21.3370
23	10.1957	11.6885	13.0905	14.8480	16.1219	17.1865	18.1373	19.0211	19.8657	20.6902	21.5095	22.3369
24	10.8564	12.4011	13.8484	15.6587	16.9686	18.0618	19.0373	19.9432	20.8084	21.6525	22.4908	23.3367
25	11.5240	13.1197	14.6114	16.4734	17.8184	18.9398	19.9393	20.8670	21.7524	22.6156	23.4724	24.3366
26	12.1981	13.8439	15.3792	17.2919	18.6714	19.8202	20.8434	21.7924	22.6975	23.5794	24.4544	25.3365
27	12.8785	14.5734	16.1514	18.1139	19.5272	20.7030	21.7494	22.7192	23.6437	24.5440	25.4367	26.3363
28	13.5647	15.3079	16.9279	18.9392	20.3857	21.5880	22.6572	23.6475	24.5909	25.5093	26.4195	27.3362
29	14.2564	16.0471	17.7084	19.7677	21.2468	22.4751	23.5666	24.5770	25.5391	26.4751	27.4025	28.3361
30	14.9534	16.7908	18.4926	20.5992	22.1103	23.3641	24.4776	25.5078	26.4881	27.4416	28.3858	29.3360
31	15.6555	17.5387	19.2806	21.4336	22.9762	24.2551	25.3901	26.4397	27.4381	28.4087	29.3694	30.3359
32	16.3622	18.2907	20.0719	22.2706	23.8442	25.1478	26.3041	27.3728	28.3889	29.3763	30.3533	31.3359
33	17.0735	19.0466	20.8665	23.1102	24.7143	26.0422	27.2194	28.3069	29.3405	30.3444	31.3375	32.3358
34	17.7891	19.8062	21.6643	23.9523	25.5864	26.9383	28.1361	29.2421	30.2928	31.3130	32.3219	33.3357
35	18.5089	20.5694	22.4650	24.7966	26.4604	27.8359	29.0540	30.1782	31.2458	32.2821	33.3065	34.3356
36	19.2327	21.3359	23.2686	25.6433	27.3362	28.7350	29.9730	31.1152	32.1995	33.2517	34.2913	35.3356
37	19.9603	22.1056	24.0749	26.4921	28.2138	29.6355	30.8933	32.0532	33.1539	34.2216	35.2764	36.3355
38	20.6914	22.8785	24.8839	27.3429	29.0931	30.5373	31.8146	32.9919	34.1089	35.1920	36.2617	37.3354
39	21.4262	23.6544	25.6954	28.1958	29.9739	31.4405	32.7369	33.9316	35.0645	36.1628	37.2472	38.3354
40	22.1643	24.4330	26.5093	29.0505	30.8563	32.3450	33.6603	34.8719	36.0207	37.1340	38.2328	39.3353
41	22.9056	25.2145	27.3256	29.9071	31.7402	33.2506	34.5846	35.8131	36.9774	38.1055	39.2187	40.3353
42	23.6501	25.9987	28.1441	30.7654	32.6255	34.1574	35.5099	36.7550	37.9347	39.0774	40.2047	41.3352
43	24.3976	26.7853	28.9647	31.6255	33.5122	35.0653	36.4361	37.6975	38.8924	40.0496	41.1909	42.3352
44	25.1480	27.5746	29.7875	32.4871	34.4002	35.9744	37.3631	38.6408	39.8507	41.0222	42.1773	43.3352
45	25.9012	28.3661	30.6122	33.3504	35.2896	36.8844	38.2910	39.5847	40.8095	41.9950	43.1638	44.3351
46	26.6572	29.1601	31.4390	34.2152	36.1801	37.7955	39.2197	40.5292	41.7687	42.9682	44.1505	45.3351
47	27.4158	29.9562	32.2676	35.0814	37.0718	38.7075	40.1492	41.4744	42.7284	43.9417	45.1373	46.3350
48	28.1770	30.7545	33.0981	35.9491	37.9648	39.6205	41.0794	42.4201	43.6885	44.9154	46.1243	47.3350
49	28.9407	31.5549	33.9303	36.8182	38.8588	40.5344	42.0104	43.3664	44.6491	45.8895	47.1114	48.3350

Tabla de Cuantiles de la Distribución Chi-Cuadrado

En el margen superior se lee $P(\chi^2 \leq x)$ para los valores de x que figuran en el cuerpo de la tabla y en el margen izquierdo los grados de libertad (ν).

ν	0.55	0.60	0.65	0.70	0.75	0.80	0.85	0.90	0.95	0.975	0.99	0.999
1	0.5707	0.7083	0.8735	1.0742	1.3233	1.6424	2.0723	2.7055	3.8415	5.0239	6.6349	10.8278
2	1.5970	1.8326	2.0996	2.4079	2.7726	3.2189	3.7942	4.6052	5.9915	7.3777	9.2103	13.8150
3	2.6430	2.9462	3.2831	3.6649	4.1083	4.6416	5.3171	6.2514	7.8147	9.3484	11.3448	16.2667
4	3.6871	4.0446	4.4377	4.8784	5.3853	5.9886	6.7449	7.7794	9.4877	11.1433	13.2767	18.4670
5	4.7278	5.1319	5.5731	6.0644	6.6257	7.2893	8.1152	9.2364	11.0705	12.8325	15.0863	20.5147
6	5.7652	6.2108	6.6948	7.2311	7.8408	8.5581	9.4461	10.6446	12.5916	14.4494	16.8118	22.4577
7	6.8000	7.2832	7.8061	8.3834	9.0371	9.8033	10.7479	12.0170	14.0672	16.0128	18.4753	24.3215
8	7.8325	8.3505	8.9094	9.5245	10.2189	11.0301	12.0271	13.3616	15.5073	17.5345	20.0902	26.1248
9	8.8632	9.4136	10.0060	10.6564	11.3887	12.2421	13.2880	14.6837	16.9190	19.0228	21.6661	27.8768
10	9.8922	10.4732	11.0971	11.7807	12.5489	13.4420	14.5339	15.9872	18.3070	20.4832	23.2093	29.5881
11	10.9199	11.5298	12.1836	12.8987	13.7007	14.6314	15.7671	17.2750	19.6751	21.9201	24.7250	31.2645
12	11.9463	12.5838	13.2661	14.0111	14.8454	15.8120	16.9893	18.5493	21.0261	23.3367	26.2170	32.9094
13	12.9717	13.6356	14.3451	15.1187	15.9839	16.9848	18.2020	19.8119	22.3620	24.7356	27.6882	34.5288
14	13.9961	14.6853	15.4209	16.2221	17.1169	18.1508	19.4062	21.0642	23.6848	26.1189	29.1412	36.1237
15	15.0197	15.7332	16.4940	17.3217	18.2451	19.3107	20.6030	22.3071	24.9958	27.4884	30.5779	37.6976
16	16.0425	16.7795	17.5646	18.4179	19.3689	20.4651	21.7931	23.5418	26.2962	28.8454	32.0000	39.2529
17	17.0646	17.8244	18.6330	19.5110	20.4887	21.6146	22.9770	24.7690	27.5871	30.1910	33.4086	40.7896
18	18.0860	18.8679	19.6993	20.6014	21.6049	22.7595	24.1555	25.9894	28.8693	31.5264	34.8053	42.3123
19	19.1069	19.9102	20.7638	21.6891	22.7178	23.9004	25.3288	27.2036	30.1435	32.8523	36.1909	43.8211
20	20.1272	20.9514	21.8265	22.7745	23.8277	25.0375	26.4976	28.4120	31.4105	34.1696	37.5662	45.3147
21	21.1470	21.9915	22.8876	23.8578	24.9348	26.1711	27.6620	29.6151	32.6706	35.4789	38.9322	46.7966
22	22.1663	23.0307	23.9473	24.9390	26.0393	27.3014	28.8225	30.8133	33.9244	36.7807	40.2893	48.2681
23	23.1852	24.0689	25.0055	26.0184	27.1413	28.4288	29.9792	32.0069	35.1725	38.0757	41.6384	49.7280
24	24.2037	25.1063	26.0625	27.0960	28.2412	29.5533	31.1325	33.1962	36.4150	39.3641	42.9798	51.1785
25	25.2218	26.1430	27.1183	28.1719	29.3388	30.6752	32.2825	34.3816	37.6525	40.6465	44.3141	52.6197
26	26.2395	27.1789	28.1730	29.2463	30.4346	31.7946	33.4295	35.5632	38.8851	41.9232	45.6418	54.0516
27	27.2569	28.2141	29.2266	30.3193	31.5284	32.9117	34.5736	36.7412	40.1133	43.1945	46.9630	55.4766
28	28.2740	29.2486	30.2791	31.3909	32.6205	34.0266	35.7150	37.9159	41.3371	44.4608	48.2783	56.8922
29	29.2908	30.2825	31.3308	32.4612	33.7109	35.1394	36.8538	39.0875	42.5570	45.7223	49.5880	58.3008
30	30.3073	31.3159	32.3815	33.5302	34.7997	36.2502	37.9902	40.2560	43.7730	46.9793	50.8921	59.7024
31	31.3235	32.3486	33.4314	34.5981	35.8871	37.3591	39.1244	41.4217	44.9854	48.2319	52.1913	61.0983
32	32.3394	33.3809	34.4804	35.6649	36.9730	38.4663	40.2563	42.5848	46.1943	49.4804	53.4859	62.4871
33	33.3551	34.4126	35.5287	36.7307	38.0575	39.5718	41.3861	43.7452	47.3999	50.7251	54.7754	63.8701
34	34.3706	35.4438	36.5763	37.7954	39.1408	40.6757	42.5140	44.9032	48.6024	51.9660	56.0610	65.2461
35	35.3858	36.4746	37.6231	38.8591	40.2228	41.7780	43.6399	46.0588	49.8018	53.2034	57.3421	66.6198
36	36.4008	37.5049	38.6693	39.9220	41.3036	42.8788	44.7641	47.2122	50.9985	54.4373	58.6192	67.9842
37	37.4156	38.5349	39.7148	40.9839	42.3833	43.9782	45.8864	48.3634	52.1923	55.6680	59.8925	69.3463
38	38.4302	39.5643	40.7597	42.0450	43.4619	45.0763	47.0072	49.5126	53.3836	56.8955	61.1620	70.7037
39	39.4446	40.5935	41.8040	43.1054	44.5395	46.1730	48.1263	50.6598	54.5722	58.1201	62.4280	72.0541
40	40.4589	41.6222	42.8477	44.1649	45.6160	47.2685	49.2439	51.8051	55.7585	59.3417	63.6908	73.4022
41	41.4729	42.6506	43.8909	45.2236	46.6916	48.3628	50.3599	52.9485	56.9424	60.5606	64.9501	74.7456
42	42.4868	43.6786	44.9335	46.2817	47.7662	49.4560	51.4746	54.0902	58.1241	61.7768	66.2063	76.0844
43	43.5005	44.7063	45.9757	47.3390	48.8400	50.5480	52.5879	55.2302	59.3035	62.9904	67.4595	77.4185
44	44.5141	45.7336	47.0173	48.3957	49.9129	51.6389	53.6998	56.3686	60.4809	64.2014	68.7095	78.7503
45	45.5274	46.7607	48.0584	49.4517	50.9849	52.7288	54.8105	57.5053	61.6562	65.4101	69.9569	80.0774
46	46.5407	47.7874	49.0991	50.5071	52.0562	53.8177	55.9199	58.6405	62.8296	66.6165	71.2014	81.3999
47	47.5538	48.8139	50.1394	51.5619	53.1267	54.9056	57.0281	59.7743	64.0011	67.8207	72.4432	82.7201
48	48.5668	49.8401	51.1792	52.6161	54.1964	55.9926	58.1352	60.9066	65.1708	69.0226	73.6827	84.0379
49	49.5796	50.8659	52.2186	53.6697	55.2653	57.0786	59.2411	62.0375	66.3386	70.2224	74.9194	85.3511

Tabla de Cuantiles de la Distribución F

Grados de libertad del numerador en extremo superior izquierdo, grados de libertad del denominador en margen izquierdo de cada fila. En el margen superior se lee $P(F \leq x)$ para los valores de x en el cuerpo de la tabla.

1	0.001	0.025	0.050	0.075	0.100	0.125	0.825	0.850	0.875	0.900	0.925	0.950	0.975	0.990
1	0.0000	0.0000	0.0062	0.0140	0.0251	0.0396	0.0576	17.349	25.274	39.863	71.384	161.44	647.79	4052.1
2	0.0000	0.0000	0.0050	0.0113	0.0202	0.0317	0.0460	5.2072	6.5333	8.5263	11.852	18.512	38.506	98.501
3	0.0000	0.0000	0.0046	0.0105	0.0187	0.0293	0.0424	3.7030	4.4651	5.5383	7.1865	10.128	17.443	34.116
4	0.0000	0.0000	0.0045	0.0100	0.0179	0.0281	0.0407	3.1620	3.7468	4.5448	5.7219	7.7086	12.217	21.197
5	0.0000	0.0000	0.0043	0.0098	0.0175	0.0274	0.0397	2.8878	3.3890	4.0604	5.0278	6.6079	10.006	16.258
6	0.0000	0.0000	0.0043	0.0096	0.0172	0.0269	0.0390	2.7231	3.1761	3.7760	4.6269	5.9874	8.8131	13.745
7	0.0000	0.0000	0.0042	0.0095	0.0170	0.0266	0.0385	2.6134	3.0354	3.5894	4.3670	5.5915	8.0727	12.246
8	0.0000	0.0000	0.0042	0.0094	0.0168	0.0264	0.0382	2.5352	2.9356	3.4579	4.1852	5.3176	7.5709	11.258
9	0.0000	0.0000	0.0042	0.0094	0.0167	0.0262	0.0379	2.4766	2.8611	3.3603	4.0510	5.1174	7.2093	10.561
10	0.0000	0.0000	0.0041	0.0093	0.0166	0.0260	0.0377	2.4312	2.8035	3.2850	3.9480	4.9646	6.9367	10.044
11	0.0000	0.0000	0.0041	0.0093	0.0165	0.0259	0.0375	2.3949	2.7576	3.2252	3.8665	4.8443	6.7241	9.6461
12	0.0000	0.0000	0.0041	0.0092	0.0165	0.0258	0.0373	2.3653	2.7202	3.1766	3.8004	4.7472	6.5538	9.3303
13	0.0000	0.0000	0.0041	0.0092	0.0164	0.0257	0.0372	2.3407	2.6891	3.1362	3.7457	4.6672	6.4143	9.0738
14	0.0000	0.0000	0.0041	0.0092	0.0164	0.0257	0.0371	2.3198	2.6628	3.1022	3.6997	4.6001	6.2979	8.8617
15	0.0000	0.0000	0.0041	0.0092	0.0163	0.0256	0.0370	2.3020	2.6404	3.0732	3.6605	4.5431	6.1995	8.6832
16	0.0000	0.0000	0.0041	0.0091	0.0163	0.0256	0.0369	2.2865	2.6210	3.0481	3.6267	4.4940	6.1151	8.5309
17	0.0000	0.0000	0.0040	0.0091	0.0163	0.0255	0.0369	2.2730	2.6040	3.0262	3.5972	4.4513	6.0420	8.3998
18	0.0000	0.0000	0.0040	0.0091	0.0162	0.0255	0.0368	2.2611	2.5891	3.0070	3.5714	4.4139	5.9781	8.2855
19	0.0000	0.0000	0.0040	0.0091	0.0162	0.0254	0.0368	2.2506	2.5758	2.9899	3.5484	4.3808	5.9216	8.1850
20	0.0000	0.0000	0.0040	0.0091	0.0162	0.0254	0.0367	2.2411	2.5640	2.9747	3.5280	4.3513	5.8715	8.0960
21	0.0000	0.0000	0.0040	0.0091	0.0162	0.0254	0.0367	2.2326	2.5533	2.9610	3.5096	4.3248	5.8266	8.0166
22	0.0000	0.0000	0.0040	0.0091	0.0162	0.0253	0.0366	2.2249	2.5437	2.9486	3.4930	4.3009	5.7863	7.9453
23	0.0000	0.0000	0.0040	0.0091	0.0161	0.0253	0.0366	2.2179	2.5350	2.9374	3.4780	4.2793	5.7498	7.8811
24	0.0000	0.0000	0.0040	0.0091	0.0161	0.0253	0.0365	2.2116	2.5270	2.9271	3.4643	4.2597	5.7166	7.8229
25	0.0000	0.0000	0.0040	0.0090	0.0161	0.0253	0.0365	2.2057	2.5197	2.9177	3.4518	4.2417	5.6864	7.7698
26	0.0000	0.0000	0.0040	0.0090	0.0161	0.0252	0.0365	2.2004	2.5130	2.9091	3.4403	4.2252	5.6586	7.7213
27	0.0000	0.0000	0.0040	0.0090	0.0161	0.0252	0.0365	2.1954	2.5068	2.9012	3.4297	4.2100	5.6331	7.6767
28	0.0000	0.0000	0.0040	0.0090	0.0161	0.0252	0.0364	2.1908	2.5011	2.8938	3.4199	4.1960	5.6096	7.6357
29	0.0000	0.0000	0.0040	0.0090	0.0161	0.0252	0.0364	2.1866	2.4958	2.8870	3.4108	4.1830	5.5878	7.5977
30	0.0000	0.0000	0.0040	0.0090	0.0161	0.0252	0.0364	2.1826	2.4908	2.8807	3.4023	4.1709	5.5675	7.5624
31	0.0000	0.0000	0.0040	0.0090	0.0161	0.0252	0.0364	2.1789	2.4862	2.8748	3.3944	4.1596	5.5487	7.5297
32	0.0000	0.0000	0.0040	0.0090	0.0160	0.0251	0.0363	2.1755	2.4819	2.8693	3.3871	4.1491	5.5311	7.4992
33	0.0000	0.0000	0.0040	0.0090	0.0160	0.0251	0.0363	2.1722	2.4778	2.8641	3.3802	4.1393	5.5147	7.4708
34	0.0000	0.0000	0.0040	0.0090	0.0160	0.0251	0.0363	2.1692	2.4741	2.8592	3.3737	4.1300	5.4993	7.4441
35	0.0000	0.0000	0.0040	0.0090	0.0160	0.0251	0.0363	2.1663	2.4705	2.8547	3.3676	4.1213	5.4848	7.4191
36	0.0000	0.0000	0.0040	0.0090	0.0160	0.0251	0.0363	2.1636	2.4671	2.8503	3.3619	4.1132	5.4712	7.3956
37	0.0000	0.0000	0.0040	0.0090	0.0160	0.0251	0.0363	2.1611	2.4639	2.8463	3.3565	4.1055	5.4584	7.3735
38	0.0000	0.0000	0.0040	0.0090	0.0160	0.0251	0.0363	2.1587	2.4609	2.8424	3.3514	4.0982	5.4463	7.3526
39	0.0000	0.0000	0.0040	0.0090	0.0160	0.0251	0.0362	2.1564	2.4581	2.8388	3.3465	4.0913	5.4348	7.3328
40	0.0000	0.0000	0.0040	0.0090	0.0160	0.0251	0.0362	2.1542	2.4554	2.8353	3.3419	4.0847	5.4239	7.3142
41	0.0000	0.0000	0.0040	0.0090	0.0160	0.0251	0.0362	2.1521	2.4528	2.8321	3.3376	4.0785	5.4137	7.2964
42	0.0000	0.0000	0.0040	0.0090	0.0160	0.0251	0.0362	2.1502	2.4504	2.8290	3.3334	4.0727	5.4039	7.2796
43	0.0000	0.0000	0.0040	0.0090	0.0160	0.0250	0.0362	2.1483	2.4481	2.8260	3.3295	4.0670	5.3946	7.2636
44	0.0000	0.0000	0.0040	0.0090	0.0160	0.0250	0.0362	2.1466	2.4459	2.8232	3.3258	4.0617	5.3857	7.2483
45	0.0000	0.0000	0.0040	0.0090	0.0160	0.0250	0.0362	2.1449	2.4437	2.8205	3.3222	4.0566	5.3773	7.2339
46	0.0000	0.0000	0.0040	0.0090	0.0160	0.0250	0.0362	2.1432	2.4417	2.8179	3.3187	4.0517	5.3692	7.2200
47	0.0000	0.0000	0.0040	0.0090	0.0160	0.0250	0.0362	2.1417	2.4398	2.8154	3.3155	4.0471	5.3615	7.2068
48	0.0000	0.0000	0.0040	0.0090	0.0160	0.0250	0.0362	2.1402	2.4380	2.8131	3.3124	4.0426	5.3541	7.1942
49	0.0000	0.0000	0.0040	0.0090	0.0160	0.0250	0.0361	2.1388	2.4362	2.8108	3.3094	4.0384	5.3471	7.1822
50	0.0000	0.0000	0.0040	0.0090	0.0160	0.0250	0.0361	2.1374	2.4345	2.8087	3.3065	4.0343	5.3403	7.1706

Tabla de Cuantiles de la Distribución F

Grados de libertad del numerador en extremo superior izquierdo, grados de libertad del denominador en margen izquierdo de cada fila. En el margen superior se lee P(F ≤ x) para los valores de x en el cuerpo de la tabla.

2	0.001	0.025	0.050	0.075	0.100	0.125	0.825	0.850	0.875	0.900	0.925	0.950	0.975	0.990
1	0.0010	0.0260	0.0540	0.0844	0.1173	0.1531	15.826	21.722	31.499	49.500	88.388	199.49	799.48	4999.3
2	0.0010	0.0256	0.0526	0.0811	0.1111	0.1429	4.7143	5.6667	7.0000	9.0000	12.333	19.000	39.000	99.000
3	0.0010	0.0255	0.0522	0.0800	0.1091	0.1397	3.2944	3.8133	4.5000	5.4624	6.9343	9.5521	16.044	30.816
4	0.0010	0.0255	0.0520	0.0795	0.1082	0.1381	2.7809	3.1640	3.6569	4.3246	5.3030	6.9443	10.649	17.999
5	0.0010	0.0254	0.0518	0.0792	0.1076	0.1372	2.5202	2.8395	3.2435	3.7797	4.5456	5.7861	8.4336	13.274
6	0.0010	0.0254	0.0517	0.0790	0.1072	0.1365	2.3634	2.6462	3.0000	3.4633	4.1138	5.1432	7.2599	10.924
7	0.0010	0.0254	0.0517	0.0788	0.1070	0.1361	2.2589	2.5183	2.8401	3.2574	3.8363	4.7374	6.5415	9.5465
8	0.0010	0.0254	0.0516	0.0787	0.1068	0.1358	2.1844	2.4274	2.7272	3.1131	3.6435	4.4590	6.0595	8.6491
9	0.0010	0.0254	0.0516	0.0786	0.1066	0.1355	2.1287	2.3597	2.6433	3.0064	3.5020	4.2565	5.7147	8.0215
10	0.0010	0.0254	0.0516	0.0786	0.1065	0.1353	2.0854	2.3072	2.5786	2.9245	3.3938	4.1028	5.4564	7.5595
11	0.0010	0.0254	0.0515	0.0785	0.1064	0.1352	2.0508	2.2654	2.5271	2.8595	3.3084	3.9823	5.2559	7.2057
12	0.0010	0.0254	0.0515	0.0785	0.1063	0.1350	2.0225	2.2313	2.4853	2.8068	3.2393	3.8853	5.0959	6.9266
13	0.0010	0.0254	0.0515	0.0784	0.1062	0.1349	1.9990	2.2030	2.4506	2.7632	3.1824	3.8056	4.9653	6.7009
14	0.0010	0.0254	0.0515	0.0784	0.1062	0.1348	1.9792	2.1791	2.4213	2.7265	3.1345	3.7389	4.8567	6.5149
15	0.0010	0.0254	0.0515	0.0784	0.1061	0.1347	1.9621	2.1586	2.3963	2.6952	3.0938	3.6823	4.7650	6.3588
16	0.0010	0.0254	0.0515	0.0783	0.1061	0.1347	1.9474	2.1410	2.3747	2.6682	3.0588	3.6337	4.6867	6.2263
17	0.0010	0.0254	0.0514	0.0783	0.1060	0.1346	1.9345	2.1255	2.3559	2.6446	3.0283	3.5915	4.6189	6.1121
18	0.0010	0.0254	0.0514	0.0783	0.1060	0.1345	1.9232	2.1119	2.3393	2.6239	3.0015	3.5546	4.5597	6.0129
19	0.0010	0.0254	0.0514	0.0783	0.1059	0.1345	1.9131	2.0998	2.3246	2.6056	2.9778	3.5219	4.5075	5.9259
20	0.0010	0.0253	0.0514	0.0783	0.1059	0.1344	1.9041	2.0890	2.3114	2.5893	2.9567	3.4928	4.4612	5.8490
21	0.0010	0.0253	0.0514	0.0783	0.1059	0.1344	1.8960	2.0793	2.2996	2.5746	2.9377	3.4668	4.4199	5.7804
22	0.0010	0.0253	0.0514	0.0782	0.1059	0.1343	1.8887	2.0705	2.2890	2.5613	2.9207	3.4434	4.3828	5.7190
23	0.0010	0.0253	0.0514	0.0782	0.1058	0.1343	1.8820	2.0626	2.2793	2.5493	2.9052	3.4221	4.3492	5.6637
24	0.0010	0.0253	0.0514	0.0782	0.1058	0.1343	1.8759	2.0553	2.2705	2.5383	2.8911	3.4028	4.3187	5.6136
25	0.0010	0.0253	0.0514	0.0782	0.1058	0.1342	1.8703	2.0487	2.2624	2.5283	2.8782	3.3852	4.2909	5.5680
26	0.0010	0.0253	0.0514	0.0782	0.1058	0.1342	1.8652	2.0425	2.2550	2.5191	2.8664	3.3690	4.2655	5.5263
27	0.0010	0.0253	0.0514	0.0782	0.1058	0.1342	1.8605	2.0369	2.2481	2.5106	2.8555	3.3541	4.2421	5.4881
28	0.0010	0.0253	0.0514	0.0782	0.1058	0.1342	1.8561	2.0317	2.2418	2.5028	2.8454	3.3404	4.2205	5.4529
29	0.0010	0.0253	0.0514	0.0782	0.1057	0.1341	1.8521	2.0268	2.2359	2.4955	2.8360	3.3277	4.2006	5.4205
30	0.0010	0.0253	0.0514	0.0782	0.1057	0.1341	1.8483	2.0223	2.2305	2.4887	2.8274	3.3158	4.1821	5.3903
31	0.0010	0.0253	0.0514	0.0782	0.1057	0.1341	1.8447	2.0181	2.2254	2.4824	2.8193	3.3048	4.1648	5.3624
32	0.0010	0.0253	0.0514	0.0782	0.1057	0.1341	1.8414	2.0142	2.2206	2.4765	2.8117	3.2945	4.1488	5.3363
33	0.0010	0.0253	0.0514	0.0781	0.1057	0.1341	1.8384	2.0105	2.2162	2.4710	2.8047	3.2849	4.1338	5.3120
34	0.0010	0.0253	0.0514	0.0781	0.1057	0.1341	1.8355	2.0070	2.2120	2.4658	2.7980	3.2759	4.1197	5.2893
35	0.0010	0.0253	0.0514	0.0781	0.1057	0.1340	1.8327	2.0038	2.2080	2.4609	2.7918	3.2674	4.1065	5.2679
36	0.0010	0.0253	0.0514	0.0781	0.1057	0.1340	1.8301	2.0007	2.2043	2.4563	2.7859	3.2594	4.0941	5.2479
37	0.0010	0.0253	0.0514	0.0781	0.1057	0.1340	1.8277	1.9978	2.2008	2.4520	2.7804	3.2519	4.0824	5.2290
38	0.0010	0.0253	0.0514	0.0781	0.1057	0.1340	1.8254	1.9951	2.1975	2.4479	2.7751	3.2448	4.0713	5.2112
39	0.0010	0.0253	0.0514	0.0781	0.1056	0.1340	1.8232	1.9925	2.1944	2.4440	2.7702	3.2381	4.0609	5.1944
40	0.0010	0.0253	0.0514	0.0781	0.1056	0.1340	1.8212	1.9900	2.1914	2.4404	2.7655	3.2317	4.0510	5.1785
41	0.0010	0.0253	0.0514	0.0781	0.1056	0.1340	1.8192	1.9877	2.1886	2.4369	2.7610	3.2257	4.0416	5.1634
42	0.0010	0.0253	0.0514	0.0781	0.1056	0.1340	1.8173	1.9855	2.1859	2.4336	2.7568	3.2199	4.0327	5.1491
43	0.0010	0.0253	0.0514	0.0781	0.1056	0.1339	1.8156	1.9833	2.1833	2.4304	2.7528	3.2145	4.0242	5.1356
44	0.0010	0.0253	0.0514	0.0781	0.1056	0.1339	1.8139	1.9813	2.1809	2.4274	2.7489	3.2093	4.0162	5.1226
45	0.0010	0.0253	0.0514	0.0781	0.1056	0.1339	1.8123	1.9794	2.1786	2.4245	2.7453	3.2043	4.0085	5.1103
46	0.0010	0.0253	0.0514	0.0781	0.1056	0.1339	1.8107	1.9776	2.1763	2.4218	2.7418	3.1996	4.0012	5.0986
47	0.0010	0.0253	0.0513	0.0781	0.1056	0.1339	1.8092	1.9758	2.1742	2.4192	2.7384	3.1951	3.9942	5.0874
48	0.0010	0.0253	0.0513	0.0781	0.1056	0.1339	1.8078	1.9741	2.1722	2.4167	2.7352	3.1907	3.9875	5.0767
49	0.0010	0.0253	0.0513	0.0781	0.1056	0.1339	1.8065	1.9725	2.1702	2.4143	2.7322	3.1866	3.9811	5.0665
50	0.0010	0.0253	0.0513	0.0781	0.1056	0.1339	1.8052	1.9710	2.1684	2.4120	2.7292	3.1826	3.9749	5.0566

Tabla de Cuantiles de la Distribución F

Grados de libertad del numerador en extremo superior izquierdo, grados de libertad del denominador en margen izquierdo de cada fila. En el margen superior se lee P(F ≤ x) para los valores de x en el cuerpo de la tabla.

3	0.001	0.025	0.050	0.075	0.100	0.125	0.150	0.850	0.875	0.900	0.925	0.950	0.975	0.990
1	0.0060	0.0573	0.0987	0.1391	0.1806	0.2240	0.2701	23.5718	34.1395	53.5933	95.6225	215.707	864.151	5403.53
2	0.0067	0.0623	0.1047	0.1442	0.1831	0.2222	0.2622	5.8258	7.1605	9.1618	12.4963	19.1642	39.1656	99.1640
3	0.0071	0.0648	0.1078	0.1473	0.1855	0.2235	0.2617	3.8209	4.4750	5.3908	6.7901	9.2766	15.4391	29.4567
4	0.0073	0.0662	0.1097	0.1492	0.1872	0.2246	0.2620	3.1236	3.5773	4.1909	5.0883	6.5914	9.9792	16.6942
5	0.0074	0.0672	0.1109	0.1505	0.1884	0.2255	0.2624	2.7764	3.1392	3.6195	4.3037	5.4094	7.7636	12.0599
6	0.0075	0.0679	0.1118	0.1515	0.1892	0.2261	0.2628	2.5699	2.8817	3.2888	3.8584	4.7571	6.5988	9.7796
7	0.0076	0.0684	0.1125	0.1522	0.1899	0.2267	0.2631	2.4334	2.7129	3.0741	3.5731	4.3468	5.8898	8.4513
8	0.0077	0.0688	0.1131	0.1528	0.1904	0.2271	0.2633	2.3366	2.5939	2.9238	3.3752	4.0662	5.4160	7.5910
9	0.0077	0.0691	0.1135	0.1532	0.1908	0.2274	0.2635	2.2644	2.5056	2.8129	3.2302	3.8625	5.0781	6.9920
10	0.0077	0.0694	0.1138	0.1536	0.1912	0.2277	0.2637	2.2086	2.4374	2.7277	3.1195	3.7083	4.8256	6.5523
11	0.0078	0.0696	0.1141	0.1539	0.1915	0.2280	0.2639	2.1640	2.3833	2.6602	3.0322	3.5874	4.6300	6.2167
12	0.0078	0.0698	0.1144	0.1542	0.1917	0.2282	0.2640	2.1278	2.3393	2.6055	2.9617	3.4903	4.4742	5.9525
13	0.0078	0.0699	0.1146	0.1544	0.1919	0.2283	0.2642	2.0976	2.3028	2.5603	2.9035	3.4105	4.3472	5.7394
14	0.0078	0.0700	0.1147	0.1546	0.1921	0.2285	0.2643	2.0722	2.2720	2.5222	2.8547	3.3439	4.2417	5.5639
15	0.0079	0.0702	0.1149	0.1547	0.1923	0.2286	0.2644	2.0504	2.2457	2.4898	2.8132	3.2874	4.1528	5.4170
16	0.0079	0.0703	0.1150	0.1549	0.1924	0.2288	0.2644	2.0316	2.2230	2.4618	2.7775	3.2389	4.0768	5.2922
17	0.0079	0.0704	0.1152	0.1550	0.1926	0.2289	0.2645	2.0152	2.2032	2.4374	2.7464	3.1968	4.0112	5.1850
18	0.0079	0.0704	0.1153	0.1552	0.1927	0.2290	0.2646	2.0007	2.1858	2.4160	2.7192	3.1599	3.9539	5.0919
19	0.0079	0.0705	0.1154	0.1553	0.1928	0.2291	0.2646	1.9878	2.1704	2.3970	2.6950	3.1274	3.9034	5.0103
20	0.0079	0.0706	0.1155	0.1554	0.1929	0.2291	0.2647	1.9764	2.1566	2.3801	2.6736	3.0984	3.8587	4.9382
21	0.0079	0.0706	0.1156	0.1554	0.1930	0.2292	0.2648	1.9660	2.1442	2.3649	2.6543	3.0725	3.8188	4.8740
22	0.0079	0.0707	0.1156	0.1555	0.1930	0.2293	0.2648	1.9567	2.1330	2.3512	2.6369	3.0491	3.7829	4.8166
23	0.0079	0.0708	0.1157	0.1556	0.1931	0.2293	0.2649	1.9482	2.1228	2.3387	2.6211	3.0280	3.7505	4.7648
24	0.0079	0.0708	0.1158	0.1557	0.1932	0.2294	0.2649	1.9405	2.1136	2.3274	2.6068	3.0088	3.7211	4.7181
25	0.0079	0.0708	0.1158	0.1557	0.1932	0.2294	0.2649	1.9334	2.1051	2.3170	2.5937	2.9912	3.6943	4.6755
26	0.0080	0.0709	0.1159	0.1558	0.1933	0.2295	0.2650	1.9269	2.0973	2.3075	2.5817	2.9752	3.6697	4.6365
27	0.0080	0.0709	0.1159	0.1559	0.1934	0.2295	0.2650	1.9209	2.0901	2.2987	2.5706	2.9603	3.6472	4.6009
28	0.0080	0.0710	0.1160	0.1559	0.1934	0.2296	0.2650	1.9154	2.0835	2.2906	2.5604	2.9467	3.6264	4.5681
29	0.0080	0.0710	0.1160	0.1559	0.1935	0.2296	0.2651	1.9102	2.0773	2.2831	2.5509	2.9340	3.6072	4.5378
30	0.0080	0.0710	0.1161	0.1560	0.1935	0.2297	0.2651	1.9054	2.0716	2.2761	2.5421	2.9223	3.5893	4.5097
31	0.0080	0.0710	0.1161	0.1560	0.1935	0.2297	0.2651	1.9010	2.0662	2.2695	2.5338	2.9113	3.5728	4.4837
32	0.0080	0.0711	0.1161	0.1561	0.1936	0.2297	0.2651	1.8968	2.0612	2.2635	2.5262	2.9011	3.5573	4.4594
33	0.0080	0.0711	0.1162	0.1561	0.1936	0.2298	0.2652	1.8929	2.0565	2.2577	2.5190	2.8916	3.5429	4.4368
34	0.0080	0.0711	0.1162	0.1561	0.1936	0.2298	0.2652	1.8892	2.0521	2.2524	2.5123	2.8826	3.5293	4.4156
35	0.0080	0.0711	0.1162	0.1562	0.1937	0.2298	0.2652	1.8857	2.0480	2.2474	2.5059	2.8742	3.5166	4.3958
36	0.0080	0.0712	0.1163	0.1562	0.1937	0.2298	0.2652	1.8825	2.0441	2.2426	2.5000	2.8663	3.5047	4.3771
37	0.0080	0.0712	0.1163	0.1562	0.1937	0.2299	0.2652	1.8794	2.0404	2.2381	2.4943	2.8588	3.4934	4.3595
38	0.0080	0.0712	0.1163	0.1563	0.1938	0.2299	0.2653	1.8765	2.0370	2.2339	2.4890	2.8517	3.4828	4.3430
39	0.0080	0.0712	0.1163	0.1563	0.1938	0.2299	0.2653	1.8737	2.0337	2.2299	2.4840	2.8451	3.4728	4.3274
40	0.0080	0.0712	0.1164	0.1563	0.1938	0.2299	0.2653	1.8711	2.0306	2.2261	2.4792	2.8387	3.4633	4.3126
41	0.0080	0.0713	0.1164	0.1563	0.1938	0.2299	0.2653	1.8686	2.0276	2.2225	2.4747	2.8327	3.4542	4.2986
42	0.0080	0.0713	0.1164	0.1564	0.1939	0.2300	0.2653	1.8662	2.0248	2.2191	2.4704	2.8271	3.4457	4.2853
43	0.0080	0.0713	0.1164	0.1564	0.1939	0.2300	0.2653	1.8640	2.0221	2.2158	2.4663	2.8216	3.4375	4.2726
44	0.0080	0.0713	0.1164	0.1564	0.1939	0.2300	0.2653	1.8619	2.0195	2.2127	2.4624	2.8165	3.4298	4.2606
45	0.0080	0.0713	0.1165	0.1564	0.1939	0.2300	0.2654	1.8598	2.0171	2.2097	2.4587	2.8115	3.4224	4.2492
46	0.0080	0.0713	0.1165	0.1564	0.1939	0.2300	0.2654	1.8579	2.0148	2.2069	2.4551	2.8068	3.4154	4.2383
47	0.0080	0.0713	0.1165	0.1565	0.1940	0.2301	0.2654	1.8560	2.0125	2.2042	2.4517	2.8024	3.4087	4.2279
48	0.0080	0.0714	0.1165	0.1565	0.1940	0.2301	0.2654	1.8542	2.0104	2.2016	2.4485	2.7981	3.4022	4.2180
49	0.0080	0.0714	0.1165	0.1565	0.1940	0.2301	0.2654	1.8525	2.0084	2.1991	2.4454	2.7940	3.3961	4.2085
50	0.0080	0.0714	0.1165	0.1565	0.1940	0.2301	0.2654	1.8508	2.0064	2.1967	2.4424	2.7900	3.3902	4.1994

Tabla de Cuantiles de la Distribución F

Grados de libertad del numerador en extremo superior izquierdo, grados de libertad del denominador en margen izquierdo de cada fila. En el margen superior se lee P(F ≤ x) para los valores de x en el cuerpo de la tabla.

4	0.001	0.025	0.050	0.075	0.100	0.125	0.150	0.850	0.875	0.900	0.925	0.950	0.975	0.990
1	0.0135	0.0818	0.1297	0.1748	0.2200	0.2669	0.3163	24.5825	35.5826	55.8330	99.5824	224.583	899.599	5624.26
2	0.0163	0.0939	0.1440	0.1886	0.2312	0.2735	0.3161	5.9065	7.2417	9.2434	12.5784	19.2467	39.2483	99.2513
3	0.0178	0.1002	0.1517	0.1965	0.2386	0.2795	0.3201	3.8166	4.4526	5.3427	6.7021	9.1172	15.1010	28.7100
4	0.0187	0.1041	0.1565	0.2016	0.2435	0.2838	0.3235	3.0916	3.5236	4.1072	4.9604	6.3882	9.6045	15.9771
5	0.0193	0.1068	0.1598	0.2051	0.2469	0.2869	0.3260	2.7309	3.0708	3.5202	4.1598	5.1922	7.3879	11.3919
6	0.0198	0.1087	0.1623	0.2077	0.2494	0.2892	0.3279	2.5164	2.8048	3.1808	3.7061	4.5337	6.2271	9.1484
7	0.0201	0.1102	0.1641	0.2096	0.2513	0.2909	0.3294	2.3746	2.6305	2.9605	3.4157	4.1203	5.5226	7.8467
8	0.0204	0.1114	0.1655	0.2112	0.2528	0.2924	0.3306	2.2740	2.5076	2.8064	3.2145	3.8379	5.0526	7.0061
9	0.0206	0.1123	0.1667	0.2124	0.2541	0.2935	0.3316	2.1989	2.4163	2.6927	3.0671	3.6331	4.7181	6.4221
10	0.0208	0.1131	0.1677	0.2134	0.2551	0.2945	0.3325	2.1408	2.3459	2.6053	2.9546	3.4780	4.4683	5.9944
11	0.0210	0.1137	0.1685	0.2143	0.2560	0.2953	0.3332	2.0945	2.2900	2.5362	2.8659	3.3567	4.2751	5.6683
12	0.0211	0.1143	0.1692	0.2151	0.2567	0.2960	0.3338	2.0568	2.2445	2.4801	2.7943	3.2592	4.1212	5.4119
13	0.0212	0.1147	0.1697	0.2157	0.2573	0.2966	0.3343	2.0254	2.2068	2.4337	2.7352	3.1791	3.9959	5.2053
14	0.0213	0.1152	0.1703	0.2162	0.2579	0.2971	0.3348	1.9989	2.1750	2.3947	2.6857	3.1122	3.8919	5.0354
15	0.0214	0.1155	0.1707	0.2167	0.2584	0.2975	0.3352	1.9763	2.1478	2.3614	2.6436	3.0556	3.8043	4.8932
16	0.0215	0.1158	0.1711	0.2172	0.2588	0.2979	0.3355	1.9567	2.1244	2.3327	2.6073	3.0069	3.7294	4.7726
17	0.0215	0.1161	0.1715	0.2175	0.2592	0.2983	0.3359	1.9396	2.1039	2.3077	2.5757	2.9647	3.6648	4.6689
18	0.0216	0.1164	0.1718	0.2179	0.2595	0.2986	0.3362	1.9245	2.0859	2.2858	2.5480	2.9277	3.6083	4.5790
19	0.0216	0.1166	0.1721	0.2182	0.2598	0.2989	0.3364	1.9111	2.0699	2.2663	2.5236	2.8951	3.5587	4.5002
20	0.0217	0.1168	0.1723	0.2185	0.2601	0.2992	0.3367	1.8991	2.0556	2.2489	2.5017	2.8661	3.5147	4.4307
21	0.0217	0.1170	0.1726	0.2187	0.2604	0.2994	0.3369	1.8884	2.0428	2.2333	2.4822	2.8401	3.4754	4.3688
22	0.0218	0.1172	0.1728	0.2190	0.2606	0.2997	0.3371	1.8787	2.0312	2.2193	2.4645	2.8167	3.4401	4.3134
23	0.0218	0.1173	0.1730	0.2192	0.2608	0.2999	0.3373	1.8698	2.0207	2.2065	2.4485	2.7955	3.4083	4.2635
24	0.0218	0.1175	0.1732	0.2194	0.2610	0.3000	0.3374	1.8618	2.0111	2.1949	2.4340	2.7763	3.3794	4.2185
25	0.0219	0.1176	0.1733	0.2196	0.2612	0.3002	0.3376	1.8544	2.0023	2.1842	2.4207	2.7587	3.3530	4.1774
26	0.0219	0.1178	0.1735	0.2197	0.2614	0.3004	0.3377	1.8476	1.9943	2.1745	2.4085	2.7426	3.3289	4.1400
27	0.0219	0.1179	0.1737	0.2199	0.2615	0.3005	0.3379	1.8414	1.9868	2.1655	2.3972	2.7278	3.3067	4.1056
28	0.0220	0.1180	0.1738	0.2200	0.2617	0.3007	0.3380	1.8356	1.9800	2.1571	2.3868	2.7141	3.2863	4.0740
29	0.0220	0.1181	0.1739	0.2202	0.2618	0.3008	0.3381	1.8302	1.9736	2.1494	2.3772	2.7014	3.2674	4.0449
30	0.0220	0.1182	0.1740	0.2203	0.2620	0.3009	0.3382	1.8252	1.9677	2.1422	2.3682	2.6896	3.2499	4.0179
31	0.0220	0.1183	0.1742	0.2204	0.2621	0.3010	0.3383	1.8205	1.9621	2.1355	2.3599	2.6787	3.2336	3.9928
32	0.0221	0.1184	0.1743	0.2205	0.2622	0.3012	0.3384	1.8162	1.9569	2.1293	2.3521	2.6684	3.2185	3.9695
33	0.0221	0.1184	0.1744	0.2206	0.2623	0.3013	0.3385	1.8121	1.9521	2.1234	2.3448	2.6589	3.2043	3.9477
34	0.0221	0.1185	0.1744	0.2207	0.2624	0.3013	0.3386	1.8082	1.9475	2.1179	2.3380	2.6499	3.1910	3.9273
35	0.0221	0.1186	0.1745	0.2208	0.2625	0.3014	0.3387	1.8046	1.9433	2.1128	2.3316	2.6415	3.1785	3.9082
36	0.0221	0.1187	0.1746	0.2209	0.2626	0.3015	0.3388	1.8012	1.9392	2.1079	2.3255	2.6335	3.1668	3.8903
37	0.0221	0.1187	0.1747	0.2210	0.2627	0.3016	0.3388	1.7980	1.9354	2.1033	2.3198	2.6261	3.1557	3.8734
38	0.0222	0.1188	0.1748	0.2211	0.2628	0.3017	0.3389	1.7950	1.9318	2.0990	2.3144	2.6190	3.1453	3.8575
39	0.0222	0.1188	0.1748	0.2212	0.2628	0.3018	0.3390	1.7921	1.9284	2.0948	2.3093	2.6123	3.1354	3.8425
40	0.0222	0.1189	0.1749	0.2213	0.2629	0.3018	0.3390	1.7894	1.9252	2.0909	2.3044	2.6060	3.1261	3.8283
41	0.0222	0.1189	0.1750	0.2213	0.2630	0.3019	0.3391	1.7868	1.9221	2.0872	2.2998	2.6000	3.1173	3.8148
42	0.0222	0.1190	0.1750	0.2214	0.2630	0.3020	0.3391	1.7843	1.9192	2.0837	2.2955	2.5943	3.1089	3.8021
43	0.0222	0.1190	0.1751	0.2215	0.2631	0.3020	0.3392	1.7820	1.9164	2.0804	2.2913	2.5888	3.1009	3.7899
44	0.0222	0.1191	0.1752	0.2215	0.2632	0.3021	0.3393	1.7797	1.9138	2.0772	2.2874	2.5837	3.0933	3.7784
45	0.0222	0.1191	0.1752	0.2216	0.2632	0.3021	0.3393	1.7776	1.9112	2.0742	2.2836	2.5787	3.0860	3.7674
46	0.0222	0.1192	0.1753	0.2216	0.2633	0.3022	0.3394	1.7756	1.9088	2.0712	2.2800	2.5740	3.0791	3.7570
47	0.0223	0.1192	0.1753	0.2217	0.2633	0.3022	0.3394	1.7736	1.9065	2.0685	2.2766	2.5695	3.0725	3.7469
48	0.0223	0.1192	0.1754	0.2217	0.2634	0.3023	0.3394	1.7717	1.9043	2.0658	2.2733	2.5652	3.0662	3.7374
49	0.0223	0.1193	0.1754	0.2218	0.2634	0.3023	0.3395	1.7699	1.9022	2.0633	2.2701	2.5611	3.0602	3.7283
50	0.0223	0.1193	0.1755	0.2218	0.2635	0.3024	0.3395	1.7682	1.9002	2.0608	2.2671	2.5572	3.0544	3.7195

Tabla de Cuantiles de la Distribución F

Grados de libertad del numerador en extremo superior izquierdo, grados de libertad del denominador en margen izquierdo de cada fila. En el margen superior se lee P(F ≤ x) para los valores de x en el cuerpo de la tabla.

5	0.001	0.025	0.050	0.075	0.100	0.125	0.150	0.850	0.875	0.900	0.925	0.950	0.975	0.990
1	0.0212	0.0999	0.1513	0.1989	0.2463	0.2951	0.3463	25.2171	36.4894	57.2400	102.071	230.160	921.835	5763.96
2	0.0269	0.1186	0.1728	0.2200	0.2646	0.3083	0.3522	5.9553	7.2907	9.2926	12.6279	19.2963	39.2984	99.3023
3	0.0301	0.1288	0.1849	0.2324	0.2763	0.3186	0.3602	3.8109	4.4354	5.3091	6.6435	9.0134	14.8848	28.2371
4	0.0322	0.1354	0.1926	0.2404	0.2841	0.3257	0.3662	3.0678	3.4859	4.0506	4.8756	6.2561	9.3645	15.5219
5	0.0336	0.1399	0.1980	0.2461	0.2896	0.3308	0.3706	2.6980	3.0232	3.4530	4.0642	5.0503	7.1464	10.9671
6	0.0347	0.1433	0.2020	0.2502	0.2937	0.3346	0.3741	2.4780	2.7514	3.1075	3.6047	4.3874	5.9875	8.7459
7	0.0355	0.1459	0.2051	0.2535	0.2969	0.3376	0.3768	2.3324	2.5732	2.8833	3.3107	3.9715	5.2852	7.4604
8	0.0362	0.1480	0.2075	0.2560	0.2995	0.3400	0.3789	2.2291	2.4474	2.7264	3.1070	3.6875	4.8173	6.6318
9	0.0367	0.1497	0.2095	0.2581	0.3015	0.3420	0.3807	2.1520	2.3540	2.6106	2.9578	3.4817	4.4844	6.0569
10	0.0372	0.1511	0.2112	0.2599	0.3033	0.3436	0.3822	2.0922	2.2820	2.5216	2.8438	3.3258	4.2361	5.6364
11	0.0375	0.1523	0.2126	0.2614	0.3047	0.3450	0.3834	2.0446	2.2247	2.4512	2.7540	3.2039	4.0440	5.3160
12	0.0379	0.1533	0.2138	0.2626	0.3060	0.3462	0.3845	2.0058	2.1781	2.3940	2.6815	3.1059	3.8911	5.0644
13	0.0381	0.1541	0.2148	0.2637	0.3071	0.3472	0.3855	1.9735	2.1394	2.3467	2.6216	3.0254	3.7667	4.8616
14	0.0384	0.1549	0.2157	0.2646	0.3080	0.3481	0.3863	1.9462	2.1068	2.3069	2.5714	2.9582	3.6634	4.6950
15	0.0386	0.1556	0.2165	0.2655	0.3088	0.3489	0.3870	1.9228	2.0790	2.2730	2.5287	2.9013	3.5764	4.5556
16	0.0388	0.1562	0.2172	0.2662	0.3096	0.3496	0.3877	1.9027	2.0549	2.2438	2.4920	2.8524	3.5021	4.4374
17	0.0390	0.1567	0.2178	0.2669	0.3102	0.3503	0.3883	1.8850	2.0339	2.2183	2.4600	2.8100	3.4379	4.3360
18	0.0391	0.1572	0.2184	0.2675	0.3108	0.3508	0.3888	1.8695	2.0154	2.1958	2.4319	2.7729	3.3820	4.2479
19	0.0393	0.1576	0.2189	0.2680	0.3114	0.3514	0.3893	1.8556	1.9990	2.1760	2.4071	2.7401	3.3327	4.1708
20	0.0394	0.1580	0.2194	0.2685	0.3119	0.3518	0.3897	1.8433	1.9844	2.1582	2.3850	2.7109	3.2891	4.1027
21	0.0395	0.1584	0.2198	0.2690	0.3123	0.3523	0.3901	1.8322	1.9712	2.1423	2.3652	2.6848	3.2501	4.0421
22	0.0396	0.1587	0.2202	0.2694	0.3127	0.3526	0.3904	1.8222	1.9593	2.1279	2.3473	2.6613	3.2151	3.9880
23	0.0397	0.1590	0.2206	0.2697	0.3131	0.3530	0.3908	1.8131	1.9485	2.1149	2.3311	2.6400	3.1835	3.9392
24	0.0398	0.1593	0.2209	0.2701	0.3134	0.3533	0.3911	1.8047	1.9387	2.1030	2.3163	2.6207	3.1548	3.8951
25	0.0399	0.1595	0.2212	0.2704	0.3137	0.3536	0.3914	1.7971	1.9296	2.0922	2.3028	2.6030	3.1287	3.8550
26	0.0399	0.1598	0.2215	0.2707	0.3140	0.3539	0.3916	1.7901	1.9214	2.0822	2.2904	2.5868	3.1048	3.8183
27	0.0400	0.1600	0.2217	0.2710	0.3143	0.3542	0.3919	1.7836	1.9137	2.0730	2.2790	2.5719	3.0828	3.7847
28	0.0401	0.1602	0.2220	0.2713	0.3146	0.3544	0.3921	1.7777	1.9067	2.0645	2.2685	2.5581	3.0626	3.7539
29	0.0402	0.1604	0.2222	0.2715	0.3148	0.3547	0.3923	1.7721	1.9001	2.0566	2.2587	2.5454	3.0438	3.7254
30	0.0402	0.1606	0.2224	0.2717	0.3151	0.3549	0.3925	1.7669	1.8940	2.0492	2.2496	2.5336	3.0265	3.6990
31	0.0403	0.1608	0.2226	0.2719	0.3153	0.3551	0.3927	1.7621	1.8883	2.0424	2.2412	2.5225	3.0103	3.6745
32	0.0403	0.1609	0.2228	0.2721	0.3155	0.3553	0.3929	1.7576	1.8830	2.0360	2.2333	2.5123	2.9953	3.6517
33	0.0404	0.1611	0.2230	0.2723	0.3157	0.3555	0.3930	1.7534	1.8780	2.0300	2.2259	2.5026	2.9812	3.6305
34	0.0404	0.1612	0.2232	0.2725	0.3158	0.3556	0.3932	1.7494	1.8733	2.0244	2.2189	2.4936	2.9680	3.6106
35	0.0405	0.1614	0.2233	0.2727	0.3160	0.3558	0.3934	1.7457	1.8689	2.0191	2.2124	2.4851	2.9557	3.5919
36	0.0405	0.1615	0.2235	0.2728	0.3162	0.3560	0.3935	1.7422	1.8648	2.0141	2.2062	2.4772	2.9440	3.5744
37	0.0405	0.1616	0.2236	0.2730	0.3163	0.3561	0.3936	1.7388	1.8608	2.0095	2.2004	2.4696	2.9331	3.5579
38	0.0406	0.1617	0.2238	0.2731	0.3165	0.3562	0.3938	1.7357	1.8571	2.0050	2.1950	2.4625	2.9227	3.5424
39	0.0406	0.1618	0.2239	0.2733	0.3166	0.3564	0.3939	1.7327	1.8536	2.0008	2.1898	2.4558	2.9130	3.5277
40	0.0406	0.1619	0.2240	0.2734	0.3167	0.3565	0.3940	1.7299	1.8503	1.9968	2.1849	2.4495	2.9037	3.5138
41	0.0407	0.1620	0.2241	0.2735	0.3168	0.3566	0.3941	1.7272	1.8472	1.9930	2.1802	2.4434	2.8950	3.5007
42	0.0407	0.1621	0.2243	0.2736	0.3170	0.3567	0.3942	1.7247	1.8442	1.9894	2.1758	2.4377	2.8866	3.4883
43	0.0407	0.1622	0.2244	0.2738	0.3171	0.3568	0.3943	1.7222	1.8413	1.9860	2.1716	2.4322	2.8787	3.4764
44	0.0408	0.1623	0.2245	0.2739	0.3172	0.3569	0.3944	1.7199	1.8386	1.9828	2.1675	2.4270	2.8712	3.4651
45	0.0408	0.1624	0.2246	0.2740	0.3173	0.3570	0.3945	1.7177	1.8360	1.9796	2.1637	2.4221	2.8640	3.4544
46	0.0408	0.1625	0.2247	0.2741	0.3174	0.3571	0.3946	1.7156	1.8335	1.9767	2.1601	2.4174	2.8572	3.4442
47	0.0408	0.1626	0.2248	0.2742	0.3175	0.3572	0.3947	1.7136	1.8311	1.9738	2.1566	2.4128	2.8506	3.4345
48	0.0409	0.1626	0.2248	0.2743	0.3176	0.3573	0.3948	1.7116	1.8288	1.9711	2.1532	2.4085	2.8444	3.4251
49	0.0409	0.1627	0.2249	0.2743	0.3177	0.3574	0.3948	1.7098	1.8267	1.9685	2.1500	2.4044	2.8384	3.4162
50	0.0409	0.1628	0.2250	0.2744	0.3178	0.3575	0.3949	1.7080	1.8246	1.9660	2.1469	2.4004	2.8326	3.4077

Tabla de Cuantiles de la Distribución F

Grados de libertad del numerador en extremo superior izquierdo, grados de libertad del denominador en margen izquierdo de cada fila. En el margen superior se lee P(F ≤ x) para los valores de x en el cuerpo de la tabla.

6	0.001	0.025	0.050	0.075	0.100	0.125	0.150	0.850	0.875	0.900	0.925	0.950	0.975	0.990
1	0.0282	0.1135	0.1670	0.2161	0.2648	0.3148	0.3672	25.6518	37.1106	58.2045	103.777	233.988	937.114	5858.95
2	0.0370	0.1377	0.1944	0.2431	0.2887	0.3333	0.3779	5.9880	7.3235	9.3255	12.6608	19.3295	39.3311	99.3314
3	0.0422	0.1515	0.2102	0.2592	0.3041	0.3470	0.3891	3.8058	4.4223	5.2847	6.6017	8.9407	14.7347	27.9106
4	0.0456	0.1606	0.2206	0.2698	0.3144	0.3565	0.3974	3.0497	3.4582	4.0097	4.8154	6.1631	9.1973	15.2068
5	0.0481	0.1670	0.2279	0.2774	0.3218	0.3634	0.4036	2.6733	2.9884	3.4045	3.9961	4.9503	6.9777	10.6722
6	0.0499	0.1718	0.2334	0.2831	0.3274	0.3687	0.4083	2.4493	2.7122	3.0546	3.5323	4.2839	5.8197	8.4660
7	0.0514	0.1756	0.2377	0.2875	0.3317	0.3728	0.4120	2.3009	2.5310	2.8274	3.2354	3.8660	5.1186	7.1914
8	0.0526	0.1786	0.2411	0.2911	0.3352	0.3761	0.4151	2.1955	2.4031	2.6683	3.0297	3.5806	4.6517	6.3707
9	0.0535	0.1810	0.2440	0.2940	0.3381	0.3788	0.4175	2.1167	2.3081	2.5509	2.8790	3.3738	4.3197	5.8018
10	0.0543	0.1831	0.2463	0.2964	0.3405	0.3811	0.4196	2.0557	2.2347	2.4606	2.7639	3.2172	4.0721	5.3858
11	0.0550	0.1849	0.2483	0.2985	0.3425	0.3831	0.4214	2.0071	2.1764	2.3891	2.6731	3.0946	3.8806	5.0692
12	0.0556	0.1864	0.2500	0.3002	0.3443	0.3847	0.4230	1.9673	2.1289	2.3310	2.5998	2.9961	3.7283	4.8205
13	0.0561	0.1877	0.2515	0.3018	0.3458	0.3862	0.4243	1.9343	2.0894	2.2830	2.5393	2.9153	3.6043	4.6203
14	0.0566	0.1888	0.2528	0.3031	0.3471	0.3874	0.4255	1.9064	2.0562	2.2426	2.4885	2.8477	3.5014	4.4558
15	0.0570	0.1898	0.2539	0.3043	0.3483	0.3886	0.4265	1.8825	2.0278	2.2081	2.4453	2.7905	3.4147	4.3183
16	0.0573	0.1907	0.2550	0.3054	0.3493	0.3896	0.4274	1.8618	2.0032	2.1783	2.4082	2.7413	3.3406	4.2016
17	0.0576	0.1915	0.2559	0.3063	0.3503	0.3905	0.4283	1.8437	1.9818	2.1524	2.3758	2.6987	3.2767	4.1015
18	0.0579	0.1922	0.2567	0.3072	0.3511	0.3913	0.4290	1.8278	1.9629	2.1296	2.3474	2.6613	3.2209	4.0146
19	0.0582	0.1929	0.2574	0.3079	0.3519	0.3920	0.4297	1.8136	1.9461	2.1094	2.3223	2.6283	3.1718	3.9386
20	0.0584	0.1935	0.2581	0.3086	0.3526	0.3927	0.4303	1.8010	1.9311	2.0913	2.2999	2.5990	3.1283	3.8714
21	0.0586	0.1940	0.2587	0.3093	0.3532	0.3933	0.4309	1.7896	1.9177	2.0751	2.2798	2.5727	3.0895	3.8117
22	0.0588	0.1945	0.2593	0.3099	0.3538	0.3938	0.4314	1.7793	1.9055	2.0605	2.2617	2.5491	3.0546	3.7583
23	0.0590	0.1950	0.2598	0.3104	0.3543	0.3943	0.4319	1.7699	1.8945	2.0472	2.2452	2.5277	3.0232	3.7102
24	0.0592	0.1954	0.2603	0.3109	0.3548	0.3948	0.4323	1.7614	1.8844	2.0351	2.2303	2.5082	2.9946	3.6667
25	0.0593	0.1958	0.2608	0.3114	0.3553	0.3953	0.4327	1.7536	1.8752	2.0241	2.2166	2.4904	2.9685	3.6272
26	0.0595	0.1962	0.2612	0.3118	0.3557	0.3957	0.4331	1.7464	1.8667	2.0139	2.2041	2.4741	2.9447	3.5911
27	0.0596	0.1965	0.2616	0.3122	0.3561	0.3960	0.4335	1.7397	1.8589	2.0045	2.1925	2.4591	2.9228	3.5580
28	0.0597	0.1968	0.2619	0.3126	0.3565	0.3964	0.4338	1.7336	1.8517	1.9959	2.1818	2.4453	2.9027	3.5276
29	0.0599	0.1971	0.2623	0.3129	0.3568	0.3967	0.4341	1.7279	1.8449	1.9878	2.1719	2.4324	2.8840	3.4995
30	0.0600	0.1974	0.2626	0.3133	0.3571	0.3970	0.4344	1.7226	1.8387	1.9803	2.1627	2.4205	2.8667	3.4735
31	0.0601	0.1977	0.2629	0.3136	0.3575	0.3973	0.4347	1.7176	1.8329	1.9734	2.1541	2.4094	2.8506	3.4493
32	0.0602	0.1979	0.2632	0.3139	0.3577	0.3976	0.4349	1.7130	1.8274	1.9668	2.1461	2.3991	2.8356	3.4269
33	0.0603	0.1982	0.2634	0.3141	0.3580	0.3979	0.4352	1.7086	1.8223	1.9607	2.1386	2.3894	2.8216	3.4059
34	0.0604	0.1984	0.2637	0.3144	0.3583	0.3981	0.4354	1.7046	1.8175	1.9550	2.1316	2.3803	2.8085	3.3863
35	0.0605	0.1986	0.2639	0.3147	0.3585	0.3984	0.4356	1.7007	1.8130	1.9496	2.1250	2.3718	2.7961	3.3679
36	0.0605	0.1988	0.2642	0.3149	0.3587	0.3986	0.4358	1.6971	1.8088	1.9445	2.1187	2.3637	2.7845	3.3507
37	0.0606	0.1990	0.2644	0.3151	0.3590	0.3988	0.4360	1.6937	1.8048	1.9398	2.1129	2.3562	2.7736	3.3344
38	0.0607	0.1992	0.2646	0.3153	0.3592	0.3990	0.4362	1.6904	1.8010	1.9352	2.1073	2.3490	2.7633	3.3191
39	0.0608	0.1993	0.2648	0.3155	0.3594	0.3992	0.4364	1.6874	1.7974	1.9309	2.1020	2.3423	2.7536	3.3047
40	0.0608	0.1995	0.2650	0.3157	0.3596	0.3994	0.4365	1.6845	1.7940	1.9269	2.0971	2.3359	2.7444	3.2910
41	0.0609	0.1997	0.2651	0.3159	0.3597	0.3995	0.4367	1.6817	1.7907	1.9230	2.0923	2.3298	2.7356	3.2781
42	0.0609	0.1998	0.2653	0.3161	0.3599	0.3997	0.4369	1.6791	1.7877	1.9194	2.0878	2.3240	2.7273	3.2658
43	0.0610	0.1999	0.2655	0.3162	0.3601	0.3999	0.4370	1.6766	1.7847	1.9159	2.0836	2.3185	2.7195	3.2541
44	0.0610	0.2001	0.2656	0.3164	0.3602	0.4000	0.4372	1.6742	1.7819	1.9125	2.0795	2.3133	2.7120	3.2430
45	0.0611	0.2002	0.2658	0.3165	0.3604	0.4002	0.4373	1.6719	1.7793	1.9093	2.0756	2.3083	2.7048	3.2325
46	0.0611	0.2003	0.2659	0.3167	0.3605	0.4003	0.4374	1.6698	1.7767	1.9063	2.0719	2.3035	2.6980	3.2224
47	0.0612	0.2005	0.2660	0.3168	0.3607	0.4004	0.4375	1.6677	1.7743	1.9034	2.0683	2.2990	2.6915	3.2128
48	0.0612	0.2006	0.2662	0.3170	0.3608	0.4006	0.4377	1.6657	1.7720	1.9006	2.0649	2.2946	2.6852	3.2036
49	0.0613	0.2007	0.2663	0.3171	0.3609	0.4007	0.4378	1.6638	1.7697	1.8980	2.0617	2.2904	2.6793	3.1948
50	0.0613	0.2008	0.2664	0.3172	0.3611	0.4008	0.4379	1.6620	1.7676	1.8954	2.0586	2.2864	2.6736	3.1864

Tabla de Cuantiles de la Distribución F

Grados de libertad del numerador en extremo superior izquierdo, grados de libertad del denominador en margen izquierdo de cada fila. En el margen superior se lee P(F ≤ x) para los valores de x en el cuerpo de la tabla.

7	0.001	0.025	0.050	0.075	0.100	0.125	0.150	0.850	0.875	0.900	0.925	0.950	0.975	0.990
1	0.0342	0.1239	0.1788	0.2290	0.2786	0.3294	0.3826	25.9679	37.5621	58.9062	105.019	236.767	948.203	5928.33
2	0.0461	0.1529	0.2111	0.2607	0.3070	0.3521	0.3971	6.0114	7.3469	9.3491	12.6845	19.3531	39.3557	99.3568
3	0.0533	0.1698	0.2301	0.2799	0.3253	0.3686	0.4109	3.8013	4.4120	5.2662	6.5705	8.8867	14.6244	27.6714
4	0.0582	0.1811	0.2427	0.2928	0.3378	0.3802	0.4211	3.0357	3.4371	3.9790	4.7704	6.0942	9.0741	14.9757
5	0.0617	0.1892	0.2518	0.3020	0.3468	0.3886	0.4287	2.6543	2.9618	3.3679	3.9451	4.8759	6.8530	10.4556
6	0.0644	0.1954	0.2587	0.3091	0.3537	0.3951	0.4346	2.4270	2.6823	3.0145	3.4779	4.2067	5.6955	8.2600
7	0.0666	0.2002	0.2641	0.3146	0.3591	0.4002	0.4393	2.2764	2.4988	2.7849	3.1788	3.7871	4.9949	6.9929
8	0.0683	0.2041	0.2684	0.3190	0.3634	0.4043	0.4431	2.1694	2.3692	2.6241	2.9714	3.5005	4.5285	6.1776
9	0.0698	0.2073	0.2720	0.3227	0.3670	0.4077	0.4462	2.0894	2.2728	2.5053	2.8194	3.2927	4.1970	5.6128
10	0.0710	0.2100	0.2750	0.3258	0.3700	0.4106	0.4489	2.0273	2.1983	2.4140	2.7033	3.1355	3.9498	5.2001
11	0.0720	0.2123	0.2775	0.3284	0.3726	0.4130	0.4511	1.9778	2.1391	2.3416	2.6118	3.0123	3.7586	4.8860
12	0.0730	0.2143	0.2797	0.3306	0.3748	0.4151	0.4531	1.9374	2.0908	2.2828	2.5377	2.9134	3.6065	4.6395
13	0.0737	0.2161	0.2817	0.3326	0.3767	0.4170	0.4548	1.9037	2.0508	2.2341	2.4766	2.8321	3.4827	4.4410
14	0.0744	0.2176	0.2833	0.3343	0.3784	0.4186	0.4563	1.8752	2.0170	2.1931	2.4254	2.7642	3.3799	4.2779
15	0.0750	0.2189	0.2848	0.3358	0.3799	0.4200	0.4576	1.8509	1.9881	2.1582	2.3818	2.7066	3.2934	4.1416
16	0.0756	0.2201	0.2862	0.3372	0.3812	0.4213	0.4588	1.8298	1.9631	2.1280	2.3442	2.6572	3.2194	4.0259
17	0.0761	0.2212	0.2874	0.3384	0.3824	0.4224	0.4598	1.8113	1.9413	2.1017	2.3115	2.6143	3.1556	3.9267
18	0.0766	0.2222	0.2884	0.3395	0.3835	0.4235	0.4608	1.7951	1.9220	2.0785	2.2828	2.5767	3.0999	3.8406
19	0.0770	0.2231	0.2894	0.3405	0.3845	0.4244	0.4617	1.7806	1.9050	2.0580	2.2574	2.5435	3.0509	3.7653
20	0.0773	0.2239	0.2903	0.3414	0.3854	0.4253	0.4625	1.7677	1.8897	2.0397	2.2347	2.5140	3.0074	3.6987
21	0.0777	0.2246	0.2911	0.3422	0.3862	0.4260	0.4632	1.7560	1.8760	2.0233	2.2144	2.4876	2.9686	3.6396
22	0.0780	0.2253	0.2919	0.3430	0.3870	0.4268	0.4639	1.7455	1.8636	2.0084	2.1961	2.4638	2.9338	3.5866
23	0.0783	0.2259	0.2926	0.3437	0.3877	0.4274	0.4645	1.7360	1.8524	1.9949	2.1795	2.4422	2.9023	3.5390
24	0.0785	0.2265	0.2932	0.3444	0.3883	0.4280	0.4650	1.7272	1.8421	1.9826	2.1643	2.4226	2.8738	3.4959
25	0.0788	0.2270	0.2938	0.3450	0.3889	0.4286	0.4656	1.7192	1.8327	1.9714	2.1505	2.4047	2.8478	3.4568
26	0.0790	0.2275	0.2944	0.3455	0.3894	0.4291	0.4661	1.7119	1.8240	1.9610	2.1378	2.3883	2.8240	3.4210
27	0.0792	0.2280	0.2949	0.3461	0.3900	0.4296	0.4665	1.7051	1.8161	1.9515	2.1261	2.3732	2.8021	3.3882
28	0.0794	0.2284	0.2954	0.3466	0.3904	0.4301	0.4670	1.6988	1.8087	1.9427	2.1153	2.3593	2.7820	3.3581
29	0.0796	0.2288	0.2958	0.3470	0.3909	0.4305	0.4674	1.6929	1.8018	1.9345	2.1052	2.3463	2.7633	3.3303
30	0.0798	0.2292	0.2962	0.3474	0.3913	0.4309	0.4677	1.6875	1.7955	1.9269	2.0959	2.3343	2.7460	3.3045
31	0.0800	0.2296	0.2966	0.3478	0.3917	0.4313	0.4681	1.6824	1.7895	1.9198	2.0872	2.3232	2.7299	3.2806
32	0.0801	0.2299	0.2970	0.3482	0.3921	0.4317	0.4684	1.6777	1.7840	1.9132	2.0791	2.3127	2.7150	3.2583
33	0.0803	0.2303	0.2974	0.3486	0.3924	0.4320	0.4688	1.6732	1.7787	1.9070	2.0715	2.3030	2.7009	3.2376
34	0.0804	0.2306	0.2977	0.3489	0.3928	0.4323	0.4690	1.6691	1.7739	1.9012	2.0644	2.2938	2.6878	3.2182
35	0.0806	0.2308	0.2980	0.3493	0.3931	0.4326	0.4693	1.6651	1.7692	1.8957	2.0577	2.2852	2.6755	3.1999
36	0.0807	0.2311	0.2983	0.3496	0.3934	0.4329	0.4696	1.6614	1.7649	1.8905	2.0514	2.2771	2.6639	3.1829
37	0.0808	0.2314	0.2986	0.3499	0.3937	0.4332	0.4699	1.6579	1.7608	1.8856	2.0454	2.2695	2.6530	3.1668
38	0.0809	0.2316	0.2989	0.3501	0.3940	0.4334	0.4701	1.6546	1.7569	1.8810	2.0398	2.2623	2.6427	3.1516
39	0.0810	0.2319	0.2991	0.3504	0.3942	0.4337	0.4703	1.6515	1.7533	1.8767	2.0344	2.2555	2.6330	3.1373
40	0.0811	0.2321	0.2994	0.3506	0.3945	0.4339	0.4706	1.6485	1.7498	1.8725	2.0294	2.2490	2.6238	3.1238
41	0.0812	0.2323	0.2996	0.3509	0.3947	0.4341	0.4708	1.6457	1.7465	1.8686	2.0246	2.2429	2.6150	3.1109
42	0.0813	0.2325	0.2998	0.3511	0.3949	0.4344	0.4710	1.6430	1.7433	1.8649	2.0200	2.2371	2.6068	3.0988
43	0.0814	0.2327	0.3000	0.3513	0.3951	0.4346	0.4712	1.6404	1.7403	1.8613	2.0157	2.2315	2.5989	3.0872
44	0.0815	0.2329	0.3002	0.3515	0.3953	0.4348	0.4713	1.6380	1.7375	1.8579	2.0116	2.2263	2.5914	3.0762
45	0.0816	0.2331	0.3004	0.3517	0.3955	0.4350	0.4715	1.6357	1.7348	1.8547	2.0076	2.2212	2.5842	3.0658
46	0.0817	0.2332	0.3006	0.3519	0.3957	0.4351	0.4717	1.6334	1.7322	1.8516	2.0039	2.2164	2.5774	3.0558
47	0.0817	0.2334	0.3008	0.3521	0.3959	0.4353	0.4718	1.6313	1.7297	1.8486	2.0003	2.2118	2.5709	3.0463
48	0.0818	0.2335	0.3010	0.3523	0.3961	0.4355	0.4720	1.6293	1.7273	1.8458	1.9968	2.2074	2.5646	3.0372
49	0.0819	0.2337	0.3011	0.3525	0.3963	0.4356	0.4722	1.6273	1.7250	1.8431	1.9935	2.2032	2.5587	3.0285
50	0.0820	0.2338	0.3013	0.3526	0.3964	0.4358	0.4723	1.6254	1.7228	1.8405	1.9904	2.1992	2.5530	3.0202

Tabla de Cuantiles de la Distribución F

Grados de libertad del numerador en extremo superior izquierdo, grados de libertad del denominador en margen izquierdo de cada fila. En el margen superior se lee P(F ≤ x) para los valores de x en el cuerpo de la tabla.

8	0.001	0.025	0.050	0.075	0.100	0.125	0.150	0.850	0.875	0.900	0.925	0.950	0.975	0.990
1	0.0393	0.1321	0.1881	0.2389	0.2892	0.3407	0.3945	26.2082	37.9055	59.4391	105.961	238.884	956.643	5980.95
2	0.0541	0.1650	0.2243	0.2745	0.3212	0.3667	0.4120	6.0290	7.3646	9.3668	12.7022	19.3709	39.3729	99.3750
3	0.0632	0.1846	0.2459	0.2963	0.3420	0.3855	0.4280	3.7976	4.4038	5.2517	6.5463	8.8452	14.5399	27.4895
4	0.0695	0.1979	0.2606	0.3111	0.3563	0.3988	0.4398	3.0245	3.4205	3.9549	4.7356	6.0410	8.9796	14.7988
5	0.0742	0.2076	0.2712	0.3219	0.3668	0.4086	0.4486	2.6391	2.9408	3.3393	3.9055	4.8183	6.7572	10.2893
6	0.0778	0.2150	0.2793	0.3301	0.3748	0.4161	0.4555	2.4093	2.6587	2.9830	3.4354	4.1468	5.5996	8.1017
7	0.0807	0.2208	0.2857	0.3365	0.3811	0.4221	0.4610	2.2570	2.4733	2.7516	3.1345	3.7257	4.8993	6.8401
8	0.0830	0.2256	0.2909	0.3418	0.3862	0.4269	0.4654	2.1486	2.3423	2.5893	2.9258	3.4381	4.4333	6.0288
9	0.0850	0.2295	0.2951	0.3461	0.3904	0.4310	0.4691	2.0675	2.2448	2.4694	2.7728	3.2296	4.1020	5.4671
10	0.0867	0.2328	0.2988	0.3498	0.3940	0.4344	0.4723	2.0046	2.1694	2.3771	2.6558	3.0717	3.8549	5.0567
11	0.0881	0.2357	0.3018	0.3529	0.3971	0.4372	0.4750	1.9543	2.1094	2.3040	2.5635	2.9480	3.6638	4.7445
12	0.0893	0.2381	0.3045	0.3556	0.3997	0.4398	0.4773	1.9133	2.0605	2.2446	2.4889	2.8486	3.5118	4.4994
13	0.0904	0.2403	0.3068	0.3580	0.4020	0.4419	0.4793	1.8791	2.0199	2.1953	2.4273	2.7669	3.3880	4.3021
14	0.0914	0.2421	0.3089	0.3600	0.4040	0.4439	0.4811	1.8502	1.9857	2.1539	2.3756	2.6987	3.2853	4.1400
15	0.0922	0.2438	0.3107	0.3619	0.4058	0.4456	0.4827	1.8254	1.9563	2.1185	2.3316	2.6408	3.1987	4.0044
16	0.0930	0.2453	0.3123	0.3635	0.4074	0.4471	0.4841	1.8040	1.9310	2.0880	2.2936	2.5911	3.1248	3.8896
17	0.0937	0.2467	0.3138	0.3650	0.4089	0.4485	0.4854	1.7852	1.9088	2.0613	2.2606	2.5480	3.0610	3.7909
18	0.0943	0.2479	0.3151	0.3663	0.4102	0.4497	0.4866	1.7686	1.8893	2.0379	2.2316	2.5102	3.0053	3.7054
19	0.0949	0.2490	0.3163	0.3676	0.4114	0.4509	0.4876	1.7539	1.8720	2.0171	2.2060	2.4768	2.9563	3.6305
20	0.0954	0.2500	0.3174	0.3687	0.4124	0.4519	0.4886	1.7408	1.8565	1.9985	2.1831	2.4471	2.9128	3.5644
21	0.0959	0.2510	0.3184	0.3697	0.4134	0.4528	0.4895	1.7289	1.8425	1.9819	2.1625	2.4205	2.8740	3.5056
22	0.0963	0.2518	0.3194	0.3706	0.4143	0.4537	0.4903	1.7182	1.8299	1.9668	2.1440	2.3965	2.8392	3.4530
23	0.0967	0.2526	0.3202	0.3715	0.4152	0.4545	0.4910	1.7084	1.8185	1.9531	2.1272	2.3748	2.8077	3.4057
24	0.0971	0.2533	0.3210	0.3723	0.4160	0.4553	0.4917	1.6995	1.8081	1.9407	2.1119	2.3551	2.7791	3.3629
25	0.0975	0.2540	0.3217	0.3730	0.4167	0.4559	0.4924	1.6914	1.7985	1.9292	2.0979	2.3371	2.7531	3.3239
26	0.0978	0.2547	0.3224	0.3737	0.4174	0.4566	0.4930	1.6839	1.7897	1.9188	2.0851	2.3205	2.7293	3.2884
27	0.0981	0.2552	0.3231	0.3743	0.4180	0.4572	0.4935	1.6770	1.7816	1.9091	2.0732	2.3053	2.7074	3.2558
28	0.0984	0.2558	0.3237	0.3749	0.4186	0.4577	0.4940	1.6705	1.7741	1.9001	2.0623	2.2913	2.6872	3.2259
29	0.0987	0.2563	0.3242	0.3755	0.4191	0.4583	0.4945	1.6646	1.7671	1.8918	2.0522	2.2782	2.6686	3.1982
30	0.0989	0.2568	0.3247	0.3760	0.4196	0.4588	0.4950	1.6590	1.7606	1.8841	2.0427	2.2662	2.6513	3.1726
31	0.0992	0.2573	0.3252	0.3765	0.4201	0.4592	0.4954	1.6539	1.7545	1.8769	2.0339	2.2549	2.6352	3.1489
32	0.0994	0.2577	0.3257	0.3770	0.4206	0.4597	0.4958	1.6490	1.7489	1.8702	2.0257	2.2444	2.6202	3.1267
33	0.0996	0.2581	0.3261	0.3774	0.4210	0.4601	0.4962	1.6445	1.7436	1.8639	2.0180	2.2346	2.6061	3.1061
34	0.0998	0.2585	0.3266	0.3779	0.4214	0.4605	0.4966	1.6402	1.7386	1.8580	2.0108	2.2253	2.5930	3.0868
35	0.1000	0.2589	0.3269	0.3783	0.4218	0.4608	0.4969	1.6362	1.7339	1.8524	2.0040	2.2167	2.5807	3.0687
36	0.1002	0.2592	0.3273	0.3786	0.4222	0.4612	0.4973	1.6324	1.7295	1.8471	1.9976	2.2085	2.5691	3.0517
37	0.1003	0.2595	0.3277	0.3790	0.4225	0.4615	0.4976	1.6288	1.7253	1.8422	1.9916	2.2008	2.5581	3.0357
38	0.1005	0.2598	0.3280	0.3793	0.4229	0.4618	0.4979	1.6254	1.7213	1.8375	1.9859	2.1936	2.5478	3.0207
39	0.1007	0.2601	0.3283	0.3797	0.4232	0.4621	0.4982	1.6222	1.7176	1.8331	1.9805	2.1867	2.5381	3.0064
40	0.1008	0.2604	0.3286	0.3800	0.4235	0.4624	0.4984	1.6192	1.7141	1.8289	1.9754	2.1802	2.5289	2.9930
41	0.1010	0.2607	0.3289	0.3803	0.4238	0.4627	0.4987	1.6163	1.7107	1.8249	1.9705	2.1740	2.5201	2.9802
42	0.1011	0.2610	0.3292	0.3806	0.4240	0.4630	0.4989	1.6136	1.7075	1.8211	1.9659	2.1681	2.5118	2.9681
43	0.1012	0.2612	0.3295	0.3808	0.4243	0.4632	0.4992	1.6109	1.7044	1.8175	1.9615	2.1625	2.5039	2.9567
44	0.1014	0.2615	0.3297	0.3811	0.4246	0.4635	0.4994	1.6084	1.7015	1.8140	1.9573	2.1572	2.4964	2.9457
45	0.1015	0.2617	0.3300	0.3813	0.4248	0.4637	0.4996	1.6061	1.6987	1.8107	1.9533	2.1521	2.4892	2.9353
46	0.1016	0.2619	0.3302	0.3816	0.4250	0.4639	0.4998	1.6038	1.6961	1.8076	1.9495	2.1473	2.4824	2.9254
47	0.1017	0.2621	0.3305	0.3818	0.4253	0.4641	0.5000	1.6016	1.6936	1.8046	1.9459	2.1427	2.4759	2.9160
48	0.1018	0.2623	0.3307	0.3820	0.4255	0.4643	0.5002	1.5995	1.6911	1.8017	1.9424	2.1382	2.4696	2.9069
49	0.1019	0.2625	0.3309	0.3822	0.4257	0.4645	0.5004	1.5975	1.6888	1.7989	1.9390	2.1340	2.4637	2.8983
50	0.1020	0.2627	0.3311	0.3824	0.4259	0.4647	0.5006	1.5956	1.6866	1.7963	1.9358	2.1299	2.4579	2.8900

Tabla de Cuantiles de la Distribución F

Grados de libertad del numerador en extremo superior izquierdo, grados de libertad del denominador en margen izquierdo de cada fila. En el margen superior se lee P(F ≤ x) para los valores de x en el cuerpo de la tabla.

9	0.001	0.025	0.050	0.075	0.100	0.125	0.150	0.850	0.875	0.900	0.925	0.950	0.975	0.990
1	0.0437	0.1387	0.1954	0.2469	0.2976	0.3495	0.4038	26.3967	38.1751	59.8575	106.701	240.543	963.279	6022.40
2	0.0610	0.1750	0.2349	0.2856	0.3326	0.3783	0.4238	6.0427	7.3783	9.3805	12.7161	19.3847	39.3866	99.3896
3	0.0719	0.1969	0.2589	0.3096	0.3555	0.3991	0.4416	3.7945	4.3971	5.2400	6.5269	8.8123	14.4730	27.3449
4	0.0796	0.2120	0.2752	0.3260	0.3714	0.4139	0.4548	3.0153	3.4070	3.9357	4.7077	5.9988	8.9046	14.6592
5	0.0854	0.2230	0.2872	0.3381	0.3831	0.4248	0.4647	2.6268	2.9239	3.3163	3.8738	4.7725	6.6810	10.1577
6	0.0899	0.2315	0.2964	0.3473	0.3920	0.4333	0.4724	2.3949	2.6396	2.9577	3.4015	4.0990	5.5234	7.9760
7	0.0935	0.2383	0.3037	0.3547	0.3992	0.4400	0.4786	2.2411	2.4526	2.7247	3.0989	3.6767	4.8232	6.7188
8	0.0964	0.2438	0.3096	0.3607	0.4050	0.4455	0.4837	2.1316	2.3204	2.5612	2.8891	3.3881	4.3572	5.9106
9	0.0990	0.2484	0.3146	0.3656	0.4098	0.4500	0.4879	2.0496	2.2220	2.4403	2.7351	3.1789	4.0260	5.3511
10	0.1011	0.2523	0.3187	0.3698	0.4139	0.4539	0.4915	1.9860	2.1459	2.3473	2.6174	3.0204	3.7790	4.9424
11	0.1029	0.2556	0.3223	0.3734	0.4173	0.4572	0.4945	1.9351	2.0852	2.2735	2.5245	2.8962	3.5879	4.6315
12	0.1045	0.2585	0.3254	0.3765	0.4204	0.4601	0.4972	1.8935	2.0358	2.2135	2.4494	2.7964	3.4358	4.3875
13	0.1059	0.2611	0.3281	0.3792	0.4230	0.4626	0.4995	1.8589	1.9947	2.1638	2.3873	2.7144	3.3120	4.1911
14	0.1071	0.2633	0.3305	0.3816	0.4253	0.4648	0.5016	1.8296	1.9600	2.1220	2.3352	2.6458	3.2093	4.0297
15	0.1082	0.2653	0.3327	0.3838	0.4274	0.4668	0.5034	1.8044	1.9303	2.0862	2.2908	2.5876	3.1227	3.8948
16	0.1092	0.2671	0.3346	0.3857	0.4293	0.4686	0.5051	1.7827	1.9046	2.0553	2.2526	2.5377	3.0488	3.7804
17	0.1101	0.2687	0.3363	0.3874	0.4309	0.4702	0.5066	1.7636	1.8822	2.0284	2.2193	2.4943	2.9849	3.6823
18	0.1110	0.2702	0.3378	0.3890	0.4325	0.4716	0.5079	1.7468	1.8624	2.0047	2.1900	2.4563	2.9291	3.5971
19	0.1117	0.2715	0.3393	0.3904	0.4338	0.4729	0.5091	1.7319	1.8448	1.9836	2.1641	2.4227	2.8801	3.5225
20	0.1124	0.2727	0.3405	0.3917	0.4351	0.4741	0.5102	1.7185	1.8291	1.9649	2.1410	2.3928	2.8365	3.4567
21	0.1130	0.2738	0.3417	0.3929	0.4363	0.4752	0.5113	1.7064	1.8150	1.9480	2.1203	2.3661	2.7977	3.3982
22	0.1136	0.2749	0.3428	0.3939	0.4373	0.4762	0.5122	1.6956	1.8022	1.9327	2.1016	2.3419	2.7628	3.3458
23	0.1141	0.2758	0.3438	0.3950	0.4383	0.4771	0.5131	1.6856	1.7906	1.9189	2.0846	2.3201	2.7313	3.2986
24	0.1146	0.2767	0.3448	0.3959	0.4392	0.4780	0.5139	1.6766	1.7800	1.9063	2.0692	2.3002	2.7027	3.2560
25	0.1151	0.2775	0.3456	0.3968	0.4401	0.4788	0.5146	1.6683	1.7703	1.8947	2.0550	2.2821	2.6766	3.2172
26	0.1155	0.2783	0.3464	0.3976	0.4408	0.4796	0.5153	1.6607	1.7614	1.8841	2.0421	2.2655	2.6528	3.1818
27	0.1159	0.2790	0.3472	0.3983	0.4416	0.4803	0.5160	1.6536	1.7531	1.8743	2.0301	2.2501	2.6309	3.1494
28	0.1163	0.2797	0.3479	0.3990	0.4423	0.4809	0.5166	1.6471	1.7455	1.8652	2.0190	2.2360	2.6106	3.1195
29	0.1167	0.2803	0.3486	0.3997	0.4429	0.4815	0.5172	1.6410	1.7384	1.8568	2.0088	2.2229	2.5919	3.0920
30	0.1170	0.2809	0.3492	0.4003	0.4435	0.4821	0.5177	1.6354	1.7318	1.8490	1.9992	2.2107	2.5746	3.0665
31	0.1173	0.2814	0.3498	0.4009	0.4441	0.4826	0.5182	1.6301	1.7256	1.8417	1.9903	2.1994	2.5585	3.0429
32	0.1176	0.2819	0.3503	0.4015	0.4446	0.4832	0.5187	1.6252	1.7199	1.8348	1.9820	2.1888	2.5434	3.0208
33	0.1179	0.2824	0.3509	0.4020	0.4451	0.4836	0.5192	1.6205	1.7145	1.8284	1.9743	2.1789	2.5294	3.0003
34	0.1181	0.2829	0.3514	0.4025	0.4456	0.4841	0.5196	1.6162	1.7094	1.8224	1.9670	2.1696	2.5162	2.9810
35	0.1184	0.2834	0.3518	0.4030	0.4461	0.4845	0.5200	1.6121	1.7047	1.8168	1.9601	2.1608	2.5039	2.9630
36	0.1186	0.2838	0.3523	0.4034	0.4465	0.4849	0.5204	1.6083	1.7002	1.8115	1.9536	2.1526	2.4922	2.9461
37	0.1189	0.2842	0.3527	0.4038	0.4469	0.4853	0.5207	1.6046	1.6959	1.8064	1.9475	2.1449	2.4813	2.9302
38	0.1191	0.2846	0.3531	0.4042	0.4473	0.4857	0.5211	1.6012	1.6919	1.8017	1.9417	2.1375	2.4710	2.9152
39	0.1193	0.2849	0.3535	0.4046	0.4477	0.4861	0.5214	1.5979	1.6881	1.7972	1.9363	2.1306	2.4612	2.9010
40	0.1195	0.2853	0.3539	0.4050	0.4480	0.4864	0.5217	1.5948	1.6845	1.7929	1.9311	2.1240	2.4519	2.8876
41	0.1197	0.2856	0.3542	0.4053	0.4484	0.4867	0.5220	1.5918	1.6810	1.7888	1.9262	2.1178	2.4432	2.8749
42	0.1199	0.2859	0.3546	0.4057	0.4487	0.4870	0.5223	1.5890	1.6778	1.7850	1.9215	2.1119	2.4348	2.8628
43	0.1200	0.2862	0.3549	0.4060	0.4490	0.4873	0.5226	1.5864	1.6747	1.7813	1.9171	2.1062	2.4269	2.8514
44	0.1202	0.2865	0.3552	0.4063	0.4493	0.4876	0.5229	1.5838	1.6717	1.7778	1.9128	2.1009	2.4194	2.8405
45	0.1204	0.2868	0.3555	0.4066	0.4496	0.4879	0.5231	1.5814	1.6689	1.7745	1.9088	2.0958	2.4122	2.8301
46	0.1205	0.2871	0.3558	0.4069	0.4499	0.4882	0.5234	1.5791	1.6662	1.7713	1.9049	2.0909	2.4054	2.8203
47	0.1207	0.2873	0.3560	0.4071	0.4501	0.4884	0.5236	1.5769	1.6636	1.7682	1.9012	2.0862	2.3988	2.8108
48	0.1208	0.2876	0.3563	0.4074	0.4504	0.4886	0.5238	1.5747	1.6611	1.7653	1.8977	2.0817	2.3925	2.8018
49	0.1209	0.2878	0.3565	0.4076	0.4506	0.4889	0.5240	1.5727	1.6588	1.7625	1.8943	2.0774	2.3866	2.7932
50	0.1211	0.2880	0.3568	0.4079	0.4509	0.4891	0.5242	1.5707	1.6565	1.7598	1.8911	2.0733	2.3808	2.7850

Tabla de Cuantiles de la Distribución F

Grados de libertad del numerador en extremo superior izquierdo, grados de libertad del denominador en margen izquierdo de cada fila. En el margen superior se lee $P(F \leq x)$ para los valores de x en el cuerpo de la tabla.

10	0.001	0.025	0.050	0.075	0.100	0.125	0.150	0.850	0.875	0.900	0.925	0.950	0.975	0.990
1	0.0475	0.1442	0.2014	0.2533	0.3044	0.3567	0.4113	26.5488	38.3925	60.1949	107.298	241.882	968.634	6055.93
2	0.0671	0.1833	0.2437	0.2947	0.3419	0.3878	0.4334	6.0537	7.3893	9.3916	12.7271	19.3959	39.3984	99.3969
3	0.0797	0.2072	0.2697	0.3206	0.3666	0.4103	0.4528	3.7918	4.3916	5.2304	6.5112	8.7855	14.4189	27.2285
4	0.0886	0.2238	0.2875	0.3385	0.3838	0.4263	0.4671	3.0077	3.3959	3.9199	4.6850	5.9644	8.8439	14.5460
5	0.0954	0.2361	0.3007	0.3516	0.3966	0.4382	0.4780	2.6165	2.9100	3.2974	3.8478	4.7351	6.6192	10.0511
6	0.1008	0.2456	0.3108	0.3618	0.4064	0.4475	0.4864	2.3830	2.6238	2.9369	3.3736	4.0600	5.4613	7.8742
7	0.1051	0.2532	0.3189	0.3699	0.4143	0.4549	0.4933	2.2279	2.4355	2.7025	3.0697	3.6365	4.7611	6.6201
8	0.1086	0.2594	0.3256	0.3765	0.4207	0.4610	0.4989	2.1174	2.3023	2.5380	2.8589	3.3472	4.2951	5.8143
9	0.1117	0.2646	0.3311	0.3821	0.4260	0.4660	0.5035	2.0347	2.2031	2.4163	2.7041	3.1373	3.9639	5.2565
10	0.1142	0.2690	0.3358	0.3867	0.4306	0.4703	0.5075	1.9704	2.1263	2.3226	2.5858	2.9782	3.7168	4.8491
11	0.1165	0.2729	0.3398	0.3907	0.4344	0.4740	0.5109	1.9190	2.0651	2.2482	2.4923	2.8536	3.5257	4.5393
12	0.1184	0.2762	0.3433	0.3942	0.4378	0.4772	0.5139	1.8770	2.0152	2.1878	2.4167	2.7534	3.3735	4.2961
13	0.1201	0.2791	0.3464	0.3973	0.4408	0.4800	0.5165	1.8419	1.9737	2.1376	2.3542	2.6710	3.2497	4.1003
14	0.1217	0.2817	0.3491	0.4000	0.4434	0.4825	0.5188	1.8123	1.9386	2.0954	2.3017	2.6022	3.1469	3.9394
15	0.1230	0.2840	0.3515	0.4024	0.4457	0.4847	0.5209	1.7868	1.9086	2.0593	2.2570	2.5437	3.0602	3.8049
16	0.1242	0.2860	0.3537	0.4046	0.4478	0.4867	0.5227	1.7648	1.8826	2.0281	2.2185	2.4935	2.9862	3.6909
17	0.1254	0.2879	0.3556	0.4065	0.4497	0.4885	0.5244	1.7455	1.8599	2.0009	2.1850	2.4499	2.9222	3.5931
18	0.1264	0.2896	0.3574	0.4083	0.4514	0.4901	0.5259	1.7285	1.8399	1.9770	2.1555	2.4117	2.8664	3.5081
19	0.1273	0.2911	0.3590	0.4099	0.4530	0.4916	0.5273	1.7133	1.8221	1.9557	2.1293	2.3779	2.8172	3.4338
20	0.1281	0.2925	0.3605	0.4114	0.4544	0.4929	0.5285	1.6997	1.8062	1.9367	2.1061	2.3479	2.7737	3.3682
21	0.1289	0.2938	0.3618	0.4127	0.4557	0.4942	0.5297	1.6875	1.7919	1.9197	2.0851	2.3210	2.7348	3.3098
22	0.1296	0.2950	0.3631	0.4139	0.4569	0.4953	0.5308	1.6765	1.7789	1.9043	2.0663	2.2967	2.6998	3.2576
23	0.1303	0.2961	0.3643	0.4151	0.4580	0.4964	0.5317	1.6664	1.7672	1.8903	2.0492	2.2747	2.6682	3.2106
24	0.1309	0.2971	0.3653	0.4162	0.4590	0.4973	0.5327	1.6572	1.7564	1.8775	2.0336	2.2547	2.6396	3.1681
25	0.1315	0.2981	0.3663	0.4172	0.4600	0.4982	0.5335	1.6488	1.7466	1.8658	2.0193	2.2365	2.6135	3.1294
26	0.1320	0.2990	0.3673	0.4181	0.4609	0.4991	0.5343	1.6411	1.7375	1.8550	2.0062	2.2197	2.5896	3.0941
27	0.1326	0.2998	0.3681	0.4189	0.4617	0.4999	0.5350	1.6339	1.7292	1.8451	1.9941	2.2043	2.5676	3.0618
28	0.1330	0.3006	0.3689	0.4198	0.4625	0.5006	0.5357	1.6273	1.7214	1.8359	1.9829	2.1900	2.5473	3.0320
29	0.1335	0.3013	0.3697	0.4205	0.4633	0.5013	0.5364	1.6211	1.7142	1.8274	1.9726	2.1768	2.5286	3.0045
30	0.1339	0.3020	0.3704	0.4212	0.4639	0.5020	0.5370	1.6154	1.7076	1.8195	1.9629	2.1646	2.5112	2.9791
31	0.1343	0.3027	0.3711	0.4219	0.4646	0.5026	0.5376	1.6100	1.7013	1.8121	1.9539	2.1532	2.4950	2.9555
32	0.1347	0.3033	0.3718	0.4225	0.4652	0.5032	0.5381	1.6050	1.6955	1.8052	1.9455	2.1425	2.4799	2.9335
33	0.1350	0.3039	0.3724	0.4231	0.4658	0.5037	0.5386	1.6003	1.6900	1.7987	1.9377	2.1325	2.4659	2.9130
34	0.1353	0.3044	0.3729	0.4237	0.4663	0.5043	0.5391	1.5959	1.6848	1.7926	1.9303	2.1231	2.4526	2.8938
35	0.1357	0.3049	0.3735	0.4242	0.4669	0.5048	0.5396	1.5917	1.6800	1.7869	1.9234	2.1143	2.4402	2.8758
36	0.1360	0.3054	0.3740	0.4248	0.4674	0.5052	0.5400	1.5878	1.6754	1.7815	1.9168	2.1061	2.4286	2.8589
37	0.1363	0.3059	0.3745	0.4252	0.4678	0.5057	0.5404	1.5841	1.6711	1.7764	1.9107	2.0982	2.4176	2.8431
38	0.1365	0.3063	0.3749	0.4257	0.4683	0.5061	0.5408	1.5806	1.6670	1.7716	1.9048	2.0909	2.4072	2.8281
39	0.1368	0.3068	0.3754	0.4261	0.4687	0.5065	0.5412	1.5773	1.6632	1.7670	1.8993	2.0839	2.3974	2.8139
40	0.1370	0.3072	0.3758	0.4266	0.4691	0.5069	0.5416	1.5741	1.6595	1.7627	1.8941	2.0773	2.3882	2.8005
41	0.1373	0.3076	0.3762	0.4270	0.4695	0.5073	0.5419	1.5711	1.6560	1.7586	1.8891	2.0710	2.3794	2.7879
42	0.1375	0.3079	0.3766	0.4274	0.4699	0.5076	0.5422	1.5683	1.6527	1.7547	1.8844	2.0650	2.3710	2.7758
43	0.1377	0.3083	0.3770	0.4277	0.4702	0.5079	0.5426	1.5655	1.6496	1.7509	1.8799	2.0593	2.3631	2.7644
44	0.1379	0.3086	0.3773	0.4281	0.4706	0.5083	0.5429	1.5629	1.6466	1.7474	1.8756	2.0539	2.3555	2.7536
45	0.1381	0.3090	0.3777	0.4284	0.4709	0.5086	0.5431	1.5605	1.6437	1.7440	1.8715	2.0487	2.3483	2.7432
46	0.1383	0.3093	0.3780	0.4287	0.4712	0.5089	0.5434	1.5581	1.6409	1.7408	1.8676	2.0438	2.3414	2.7334
47	0.1385	0.3096	0.3783	0.4291	0.4715	0.5092	0.5437	1.5558	1.6383	1.7377	1.8638	2.0391	2.3348	2.7240
48	0.1387	0.3099	0.3786	0.4294	0.4718	0.5094	0.5440	1.5537	1.6358	1.7347	1.8603	2.0346	2.3286	2.7150
49	0.1389	0.3102	0.3789	0.4296	0.4721	0.5097	0.5442	1.5516	1.6334	1.7319	1.8568	2.0303	2.3226	2.7064
50	0.1390	0.3104	0.3792	0.4299	0.4724	0.5100	0.5444	1.5496	1.6311	1.7291	1.8536	2.0261	2.3168	2.6981

Tabla de Cuantiles de la Distribución F

Grados de libertad del numerador en extremo superior izquierdo, grados de libertad del denominador en margen izquierdo de cada fila. En el margen superior se lee P(F ≤ x) para los valores de x en el cuerpo de la tabla.

11	0.001	0.025	0.050	0.075	0.100	0.125	0.150	0.850	0.875	0.900	0.925	0.950	0.975	0.990
1	0.0508	0.1487	0.2064	0.2586	0.3101	0.3626	0.4175	26.6739	38.5712	60.4728	107.790	242.981	973.028	6083.40
2	0.0724	0.1903	0.2511	0.3023	0.3497	0.3957	0.4414	6.0627	7.3983	9.4006	12.7361	19.4050	39.4066	99.4078
3	0.0865	0.2160	0.2788	0.3298	0.3759	0.4196	0.4621	3.7895	4.3869	5.2224	6.4981	8.7633	14.3741	27.1320
4	0.0966	0.2339	0.2979	0.3489	0.3943	0.4367	0.4774	3.0014	3.3867	3.9067	4.6660	5.9358	8.7936	14.4523
5	0.1044	0.2473	0.3121	0.3631	0.4080	0.4495	0.4891	2.6079	2.8983	3.2816	3.8262	4.7040	6.5678	9.9626
6	0.1105	0.2577	0.3231	0.3741	0.4186	0.4595	0.4982	2.3729	2.6105	2.9195	3.3503	4.0274	5.4098	7.7896
7	0.1155	0.2661	0.3320	0.3829	0.4271	0.4675	0.5056	2.2167	2.4211	2.6839	3.0453	3.6030	4.7095	6.5381
8	0.1197	0.2729	0.3392	0.3901	0.4340	0.4741	0.5117	2.1054	2.2870	2.5186	2.8336	3.3129	4.2434	5.7343
9	0.1232	0.2787	0.3453	0.3961	0.4398	0.4796	0.5168	2.0220	2.1871	2.3961	2.6781	3.1025	3.9121	5.1779
10	0.1262	0.2836	0.3504	0.4012	0.4448	0.4842	0.5211	1.9572	2.1097	2.3018	2.5592	2.9430	3.6649	4.7716
11	0.1288	0.2879	0.3549	0.4056	0.4490	0.4883	0.5248	1.9054	2.0481	2.2269	2.4652	2.8179	3.4737	4.4624
12	0.1311	0.2916	0.3587	0.4095	0.4527	0.4918	0.5281	1.8629	1.9977	2.1660	2.3892	2.7173	3.3215	4.2198
13	0.1332	0.2948	0.3621	0.4128	0.4560	0.4948	0.5309	1.8275	1.9559	2.1155	2.3264	2.6346	3.1975	4.0245
14	0.1350	0.2977	0.3651	0.4158	0.4589	0.4976	0.5335	1.7975	1.9205	2.0730	2.2736	2.5655	3.0946	3.8640
15	0.1366	0.3003	0.3678	0.4185	0.4614	0.5000	0.5357	1.7718	1.8902	2.0366	2.2286	2.5068	3.0078	3.7299
16	0.1381	0.3026	0.3702	0.4209	0.4638	0.5022	0.5378	1.7496	1.8639	2.0051	2.1898	2.4564	2.9337	3.6162
17	0.1394	0.3047	0.3724	0.4230	0.4658	0.5042	0.5396	1.7300	1.8410	1.9777	2.1560	2.4126	2.8696	3.5185
18	0.1406	0.3066	0.3744	0.4250	0.4677	0.5059	0.5413	1.7128	1.8208	1.9535	2.1263	2.3742	2.8137	3.4338
19	0.1417	0.3084	0.3762	0.4268	0.4694	0.5076	0.5428	1.6975	1.8028	1.9321	2.1000	2.3402	2.7645	3.3596
20	0.1427	0.3100	0.3779	0.4284	0.4710	0.5091	0.5442	1.6837	1.7867	1.9129	2.0765	2.3100	2.7209	3.2941
21	0.1436	0.3114	0.3794	0.4299	0.4725	0.5104	0.5455	1.6714	1.7722	1.8956	2.0554	2.2829	2.6819	3.2359
22	0.1445	0.3128	0.3808	0.4313	0.4738	0.5117	0.5466	1.6602	1.7591	1.8801	2.0364	2.2585	2.6469	3.1837
23	0.1453	0.3140	0.3821	0.4326	0.4750	0.5129	0.5477	1.6500	1.7472	1.8659	2.0191	2.2364	2.6152	3.1368
24	0.1460	0.3152	0.3833	0.4338	0.4762	0.5140	0.5487	1.6407	1.7363	1.8530	2.0034	2.2163	2.5865	3.0944
25	0.1467	0.3163	0.3844	0.4349	0.4773	0.5150	0.5497	1.6321	1.7264	1.8412	1.9890	2.1979	2.5603	3.0558
26	0.1474	0.3173	0.3855	0.4359	0.4783	0.5159	0.5506	1.6243	1.7172	1.8303	1.9757	2.1811	2.5363	3.0205
27	0.1480	0.3183	0.3864	0.4369	0.4792	0.5168	0.5514	1.6170	1.7087	1.8203	1.9636	2.1655	2.5143	2.9882
28	0.1486	0.3191	0.3874	0.4378	0.4801	0.5176	0.5522	1.6103	1.7009	1.8110	1.9523	2.1512	2.4940	2.9585
29	0.1491	0.3200	0.3882	0.4386	0.4809	0.5184	0.5529	1.6040	1.6936	1.8024	1.9418	2.1379	2.4752	2.9311
30	0.1496	0.3208	0.3890	0.4394	0.4816	0.5191	0.5536	1.5982	1.6868	1.7944	1.9321	2.1256	2.4578	2.9057
31	0.1501	0.3215	0.3898	0.4402	0.4824	0.5198	0.5542	1.5928	1.6805	1.7869	1.9230	2.1141	2.4415	2.8821
32	0.1505	0.3222	0.3905	0.4409	0.4831	0.5205	0.5548	1.5877	1.6746	1.7799	1.9146	2.1033	2.4264	2.8602
33	0.1510	0.3229	0.3912	0.4416	0.4837	0.5211	0.5554	1.5829	1.6690	1.7733	1.9066	2.0933	2.4123	2.8397
34	0.1514	0.3235	0.3918	0.4422	0.4843	0.5217	0.5559	1.5784	1.6638	1.7672	1.8992	2.0838	2.3990	2.8205
35	0.1518	0.3241	0.3925	0.4428	0.4849	0.5222	0.5564	1.5742	1.6589	1.7614	1.8922	2.0750	2.3866	2.8026
36	0.1521	0.3246	0.3930	0.4434	0.4855	0.5227	0.5569	1.5702	1.6543	1.7559	1.8855	2.0666	2.3749	2.7857
37	0.1525	0.3252	0.3936	0.4439	0.4860	0.5232	0.5574	1.5665	1.6499	1.7508	1.8793	2.0587	2.3639	2.7698
38	0.1528	0.3257	0.3941	0.4444	0.4865	0.5237	0.5578	1.5629	1.6458	1.7459	1.8734	2.0513	2.3535	2.7549
39	0.1531	0.3262	0.3946	0.4449	0.4870	0.5242	0.5583	1.5595	1.6419	1.7413	1.8678	2.0443	2.3436	2.7407
40	0.1534	0.3267	0.3951	0.4454	0.4874	0.5246	0.5587	1.5563	1.6381	1.7369	1.8625	2.0376	2.3343	2.7273
41	0.1537	0.3271	0.3956	0.4459	0.4879	0.5250	0.5591	1.5533	1.6346	1.7327	1.8575	2.0312	2.3255	2.7147
42	0.1540	0.3275	0.3960	0.4463	0.4883	0.5254	0.5594	1.5504	1.6312	1.7288	1.8527	2.0252	2.3171	2.7027
43	0.1543	0.3279	0.3964	0.4467	0.4887	0.5258	0.5598	1.5476	1.6280	1.7250	1.8482	2.0195	2.3091	2.6913
44	0.1545	0.3283	0.3968	0.4471	0.4891	0.5262	0.5601	1.5450	1.6250	1.7214	1.8439	2.0140	2.3015	2.6804
45	0.1548	0.3287	0.3972	0.4475	0.4894	0.5265	0.5604	1.5425	1.6221	1.7180	1.8397	2.0088	2.2943	2.6701
46	0.1550	0.3291	0.3976	0.4479	0.4898	0.5268	0.5607	1.5401	1.6193	1.7147	1.8358	2.0039	2.2874	2.6602
47	0.1552	0.3294	0.3980	0.4482	0.4901	0.5272	0.5610	1.5378	1.6166	1.7115	1.8320	1.9991	2.2808	2.6508
48	0.1554	0.3297	0.3983	0.4486	0.4904	0.5275	0.5613	1.5356	1.6141	1.7085	1.8284	1.9946	2.2745	2.6419
49	0.1556	0.3301	0.3986	0.4489	0.4908	0.5278	0.5616	1.5334	1.6116	1.7057	1.8249	1.9902	2.2684	2.6332
50	0.1559	0.3304	0.3989	0.4492	0.4911	0.5281	0.5619	1.5314	1.6093	1.7029	1.8216	1.9861	2.2627	2.6250

Tabla de Cuantiles de la Distribución F

Grados de libertad del numerador en extremo superior izquierdo, grados de libertad del denominador en margen izquierdo de cada fila. En el margen superior se lee $P(F \leq x)$ para los valores de x en el cuerpo de la tabla.

12	0.001	0.025	0.050	0.075	0.100	0.125	0.150	0.850	0.875	0.900	0.925	0.950	0.975	0.990
1	0.0536	0.1526	0.2106	0.2631	0.3148	0.3676	0.4228	26.7785	38.7208	60.7051	108.202	243.905	976.725	6106.68
2	0.0771	0.1962	0.2574	0.3087	0.3563	0.4024	0.4482	6.0702	7.4058	9.4082	12.7437	19.4125	39.4148	99.4187
3	0.0926	0.2235	0.2865	0.3376	0.3838	0.4275	0.4700	3.7874	4.3829	5.2156	6.4870	8.7447	14.3366	27.0520
4	0.1038	0.2426	0.3068	0.3579	0.4032	0.4455	0.4862	2.9959	3.3788	3.8955	4.6501	5.9117	8.7512	14.3737
5	0.1125	0.2570	0.3220	0.3729	0.4177	0.4591	0.4986	2.6006	2.8883	3.2682	3.8079	4.6777	6.5245	9.8883
6	0.1193	0.2682	0.3338	0.3846	0.4290	0.4697	0.5083	2.3643	2.5992	2.9047	3.3306	3.9999	5.3662	7.7183
7	0.1250	0.2773	0.3432	0.3941	0.4381	0.4783	0.5162	2.2072	2.4089	2.6681	3.0246	3.5747	4.6658	6.4691
8	0.1297	0.2848	0.3511	0.4018	0.4455	0.4853	0.5227	2.0952	2.2740	2.5020	2.8121	3.2839	4.1997	5.6667
9	0.1337	0.2910	0.3576	0.4083	0.4518	0.4912	0.5281	2.0112	2.1735	2.3789	2.6560	3.0729	3.8682	5.1115
10	0.1371	0.2964	0.3632	0.4138	0.4571	0.4962	0.5328	1.9459	2.0956	2.2841	2.5366	2.9130	3.6210	4.7058
11	0.1401	0.3011	0.3680	0.4186	0.4617	0.5006	0.5368	1.8936	2.0335	2.2087	2.4422	2.7876	3.4296	4.3974
12	0.1428	0.3051	0.3722	0.4227	0.4657	0.5043	0.5403	1.8508	1.9828	2.1474	2.3657	2.6866	3.2773	4.1553
13	0.1451	0.3087	0.3759	0.4264	0.4692	0.5077	0.5434	1.8151	1.9406	2.0966	2.3026	2.6037	3.1532	3.9603
14	0.1472	0.3119	0.3792	0.4296	0.4723	0.5106	0.5461	1.7849	1.9049	2.0537	2.2495	2.5342	3.0502	3.8002
15	0.1490	0.3147	0.3821	0.4325	0.4751	0.5133	0.5486	1.7589	1.8743	2.0171	2.2042	2.4753	2.9633	3.6662
16	0.1507	0.3173	0.3848	0.4351	0.4776	0.5156	0.5508	1.7364	1.8479	1.9854	2.1652	2.4247	2.8891	3.5527
17	0.1523	0.3196	0.3872	0.4374	0.4799	0.5178	0.5528	1.7167	1.8247	1.9577	2.1312	2.3807	2.8249	3.4552
18	0.1537	0.3217	0.3893	0.4396	0.4819	0.5197	0.5546	1.6993	1.8043	1.9333	2.1013	2.3421	2.7689	3.3706
19	0.1550	0.3237	0.3913	0.4415	0.4838	0.5215	0.5562	1.6838	1.7861	1.9117	2.0747	2.3080	2.7196	3.2965
20	0.1561	0.3254	0.3931	0.4433	0.4855	0.5231	0.5577	1.6699	1.7699	1.8924	2.0511	2.2776	2.6758	3.2311
21	0.1572	0.3271	0.3948	0.4450	0.4871	0.5246	0.5591	1.6574	1.7552	1.8750	2.0299	2.2504	2.6368	3.1729
22	0.1582	0.3286	0.3964	0.4465	0.4886	0.5260	0.5604	1.6460	1.7420	1.8593	2.0107	2.2258	2.6017	3.1209
23	0.1592	0.3300	0.3978	0.4479	0.4899	0.5273	0.5616	1.6357	1.7300	1.8450	1.9933	2.2036	2.5699	3.0740
24	0.1600	0.3313	0.3991	0.4492	0.4912	0.5285	0.5627	1.6263	1.7190	1.8319	1.9774	2.1834	2.5411	3.0316
25	0.1608	0.3325	0.4004	0.4504	0.4923	0.5296	0.5637	1.6177	1.7089	1.8200	1.9629	2.1649	2.5149	2.9931
26	0.1616	0.3336	0.4015	0.4515	0.4934	0.5306	0.5647	1.6097	1.6996	1.8090	1.9496	2.1479	2.4909	2.9578
27	0.1623	0.3347	0.4026	0.4526	0.4944	0.5316	0.5656	1.6023	1.6911	1.7989	1.9373	2.1323	2.4688	2.9256
28	0.1630	0.3357	0.4036	0.4536	0.4954	0.5325	0.5665	1.5955	1.6831	1.7895	1.9259	2.1179	2.4484	2.8959
29	0.1636	0.3366	0.4046	0.4545	0.4963	0.5333	0.5672	1.5892	1.6758	1.7808	1.9154	2.1045	2.4295	2.8685
30	0.1642	0.3375	0.4055	0.4554	0.4971	0.5341	0.5680	1.5833	1.6689	1.7727	1.9056	2.0921	2.4120	2.8431
31	0.1648	0.3383	0.4063	0.4563	0.4979	0.5349	0.5687	1.5778	1.6625	1.7651	1.8964	2.0805	2.3958	2.8195
32	0.1653	0.3391	0.4071	0.4570	0.4987	0.5356	0.5694	1.5726	1.6565	1.7581	1.8879	2.0697	2.3806	2.7976
33	0.1658	0.3398	0.4079	0.4578	0.4994	0.5363	0.5700	1.5678	1.6509	1.7514	1.8798	2.0595	2.3664	2.7771
34	0.1663	0.3405	0.4086	0.4585	0.5001	0.5369	0.5706	1.5633	1.6456	1.7452	1.8723	2.0500	2.3531	2.7580
35	0.1667	0.3412	0.4093	0.4592	0.5007	0.5375	0.5711	1.5590	1.6406	1.7394	1.8652	2.0411	2.3406	2.7400
36	0.1672	0.3418	0.4099	0.4598	0.5013	0.5381	0.5717	1.5549	1.6359	1.7338	1.8586	2.0327	2.3289	2.7232
37	0.1676	0.3424	0.4105	0.4604	0.5019	0.5386	0.5722	1.5511	1.6315	1.7286	1.8523	2.0248	2.3178	2.7073
38	0.1680	0.3430	0.4111	0.4610	0.5025	0.5391	0.5727	1.5475	1.6273	1.7237	1.8463	2.0173	2.3074	2.6923
39	0.1683	0.3436	0.4117	0.4615	0.5030	0.5396	0.5731	1.5441	1.6233	1.7190	1.8407	2.0102	2.2975	2.6782
40	0.1687	0.3441	0.4122	0.4620	0.5035	0.5401	0.5736	1.5408	1.6196	1.7146	1.8353	2.0035	2.2882	2.6648
41	0.1690	0.3446	0.4127	0.4625	0.5040	0.5406	0.5740	1.5377	1.6160	1.7103	1.8303	1.9971	2.2793	2.6522
42	0.1694	0.3451	0.4132	0.4630	0.5044	0.5410	0.5744	1.5348	1.6126	1.7063	1.8254	1.9910	2.2709	2.6401
43	0.1697	0.3455	0.4137	0.4635	0.5049	0.5414	0.5748	1.5320	1.6093	1.7025	1.8209	1.9852	2.2629	2.6288
44	0.1700	0.3460	0.4141	0.4639	0.5053	0.5418	0.5752	1.5293	1.6062	1.6989	1.8165	1.9797	2.2552	2.6179
45	0.1703	0.3464	0.4146	0.4643	0.5057	0.5422	0.5755	1.5268	1.6033	1.6954	1.8123	1.9745	2.2480	2.6076
46	0.1705	0.3468	0.4150	0.4647	0.5061	0.5426	0.5759	1.5243	1.6005	1.6921	1.8083	1.9695	2.2410	2.5977
47	0.1708	0.3472	0.4154	0.4651	0.5065	0.5429	0.5762	1.5220	1.5978	1.6889	1.8045	1.9647	2.2344	2.5883
48	0.1711	0.3476	0.4158	0.4655	0.5068	0.5433	0.5765	1.5197	1.5952	1.6859	1.8008	1.9601	2.2281	2.5793
49	0.1713	0.3479	0.4161	0.4659	0.5072	0.5436	0.5768	1.5176	1.5927	1.6830	1.7973	1.9557	2.2220	2.5707
50	0.1716	0.3483	0.4165	0.4662	0.5075	0.5439	0.5771	1.5155	1.5903	1.6802	1.7940	1.9515	2.2162	2.5625

Tabla de Cuantiles de la Distribución F

Grados de libertad del numerador en extremo superior izquierdo, grados de libertad del denominador en margen izquierdo de cada fila. En el margen superior se lee $P(F \leq x)$ para los valores de x en el cuerpo de la tabla.

13	0.001	0.025	0.050	0.075	0.100	0.125	0.150	0.850	0.875	0.900	0.925	0.950	0.975	0.990
1	0.0561	0.1559	0.2143	0.2670	0.3189	0.3719	0.4272	26.8676	38.8482	60.9025	108.552	244.690	979.839	6125.77
2	0.0812	0.2014	0.2628	0.3142	0.3619	0.4081	0.4539	6.0765	7.4122	9.4145	12.7501	19.4188	39.4211	99.4223
3	0.0980	0.2300	0.2932	0.3444	0.3906	0.4343	0.4767	3.7857	4.3794	5.2098	6.4775	8.7286	14.3045	26.9829
4	0.1102	0.2503	0.3146	0.3656	0.4109	0.4532	0.4937	2.9912	3.3720	3.8859	4.6364	5.8911	8.7150	14.3064
5	0.1197	0.2655	0.3305	0.3814	0.4261	0.4674	0.5067	2.5942	2.8798	3.2567	3.7922	4.6552	6.4876	9.8248
6	0.1273	0.2774	0.3430	0.3938	0.4380	0.4786	0.5170	2.3568	2.5895	2.8920	3.3137	3.9764	5.3290	7.6575
7	0.1335	0.2871	0.3531	0.4038	0.4476	0.4876	0.5253	2.1990	2.3983	2.6545	3.0068	3.5503	4.6285	6.4100
8	0.1388	0.2952	0.3614	0.4120	0.4555	0.4951	0.5322	2.0863	2.2628	2.4876	2.7936	3.2590	4.1622	5.6089
9	0.1432	0.3019	0.3684	0.4189	0.4621	0.5013	0.5380	2.0018	2.1617	2.3640	2.6370	3.0475	3.8306	5.0545
10	0.1471	0.3077	0.3744	0.4248	0.4678	0.5067	0.5429	1.9361	2.0833	2.2687	2.5170	2.8872	3.5832	4.6496
11	0.1504	0.3127	0.3796	0.4299	0.4727	0.5113	0.5472	1.8834	2.0208	2.1930	2.4223	2.7614	3.3917	4.3416
12	0.1534	0.3171	0.3841	0.4343	0.4770	0.5153	0.5509	1.8403	1.9698	2.1313	2.3455	2.6602	3.2393	4.0998
13	0.1560	0.3210	0.3881	0.4382	0.4807	0.5189	0.5542	1.8043	1.9273	2.0802	2.2820	2.5769	3.1150	3.9052
14	0.1584	0.3245	0.3916	0.4417	0.4841	0.5220	0.5572	1.7738	1.8913	2.0370	2.2286	2.5073	3.0119	3.7452
15	0.1605	0.3276	0.3948	0.4448	0.4871	0.5248	0.5598	1.7476	1.8605	2.0001	2.1831	2.4481	2.9249	3.6115
16	0.1624	0.3304	0.3976	0.4476	0.4897	0.5274	0.5621	1.7249	1.8339	1.9682	2.1439	2.3973	2.8506	3.4981
17	0.1641	0.3329	0.4002	0.4501	0.4922	0.5297	0.5643	1.7050	1.8105	1.9404	2.1096	2.3531	2.7863	3.4007
18	0.1657	0.3352	0.4026	0.4524	0.4944	0.5318	0.5662	1.6874	1.7899	1.9158	2.0796	2.3143	2.7302	3.3162
19	0.1672	0.3373	0.4047	0.4545	0.4964	0.5337	0.5680	1.6718	1.7716	1.8940	2.0529	2.2800	2.6808	3.2422
20	0.1685	0.3393	0.4067	0.4565	0.4983	0.5354	0.5696	1.6578	1.7552	1.8745	2.0291	2.2495	2.6369	3.1769
21	0.1698	0.3410	0.4085	0.4583	0.5000	0.5370	0.5711	1.6451	1.7404	1.8570	2.0077	2.2222	2.5978	3.1187
22	0.1709	0.3427	0.4102	0.4599	0.5015	0.5385	0.5725	1.6337	1.7271	1.8411	1.9884	2.1975	2.5626	3.0667
23	0.1720	0.3442	0.4117	0.4614	0.5030	0.5399	0.5738	1.6233	1.7149	1.8267	1.9709	2.1752	2.5308	3.0199
24	0.1730	0.3456	0.4132	0.4628	0.5044	0.5412	0.5750	1.6137	1.7038	1.8136	1.9549	2.1548	2.5019	2.9775
25	0.1739	0.3470	0.4145	0.4642	0.5056	0.5424	0.5761	1.6050	1.6936	1.8015	1.9403	2.1362	2.4756	2.9389
26	0.1748	0.3482	0.4158	0.4654	0.5068	0.5435	0.5771	1.5969	1.6843	1.7904	1.9268	2.1192	2.4515	2.9038
27	0.1756	0.3494	0.4170	0.4665	0.5079	0.5445	0.5781	1.5895	1.6756	1.7802	1.9144	2.1034	2.4293	2.8715
28	0.1764	0.3505	0.4181	0.4676	0.5089	0.5455	0.5790	1.5826	1.6676	1.7708	1.9030	2.0889	2.4089	2.8418
29	0.1771	0.3515	0.4191	0.4686	0.5099	0.5464	0.5799	1.5762	1.6601	1.7620	1.8923	2.0755	2.3900	2.8144
30	0.1777	0.3525	0.4201	0.4696	0.5108	0.5473	0.5807	1.5702	1.6532	1.7538	1.8824	2.0630	2.3724	2.7890
31	0.1784	0.3534	0.4210	0.4705	0.5117	0.5481	0.5814	1.5646	1.6467	1.7461	1.8732	2.0513	2.3561	2.7655
32	0.1790	0.3542	0.4219	0.4714	0.5125	0.5489	0.5822	1.5594	1.6407	1.7390	1.8646	2.0404	2.3409	2.7435
33	0.1796	0.3550	0.4227	0.4722	0.5133	0.5496	0.5828	1.5545	1.6350	1.7323	1.8565	2.0302	2.3266	2.7231
34	0.1801	0.3558	0.4235	0.4729	0.5140	0.5503	0.5835	1.5499	1.6296	1.7260	1.8489	2.0207	2.3133	2.7039
35	0.1806	0.3566	0.4243	0.4737	0.5147	0.5510	0.5841	1.5456	1.6246	1.7201	1.8418	2.0117	2.3008	2.6859
36	0.1811	0.3573	0.4250	0.4744	0.5154	0.5516	0.5847	1.5415	1.6199	1.7145	1.8351	2.0032	2.2890	2.6691
37	0.1816	0.3579	0.4257	0.4750	0.5160	0.5522	0.5852	1.5376	1.6154	1.7092	1.8287	1.9952	2.2779	2.6532
38	0.1821	0.3586	0.4263	0.4756	0.5166	0.5528	0.5858	1.5340	1.6111	1.7042	1.8227	1.9877	2.2674	2.6382
39	0.1825	0.3592	0.4269	0.4762	0.5172	0.5533	0.5863	1.5305	1.6071	1.6995	1.8170	1.9805	2.2575	2.6241
40	0.1829	0.3598	0.4275	0.4768	0.5177	0.5538	0.5867	1.5272	1.6033	1.6950	1.8116	1.9738	2.2481	2.6107
41	0.1833	0.3603	0.4281	0.4774	0.5183	0.5543	0.5872	1.5241	1.5997	1.6908	1.8065	1.9673	2.2392	2.5981
42	0.1837	0.3608	0.4286	0.4779	0.5188	0.5548	0.5876	1.5211	1.5962	1.6867	1.8016	1.9612	2.2307	2.5860
43	0.1841	0.3613	0.4291	0.4784	0.5192	0.5552	0.5881	1.5182	1.5929	1.6829	1.7970	1.9554	2.2227	2.5746
44	0.1844	0.3618	0.4296	0.4789	0.5197	0.5557	0.5885	1.5155	1.5898	1.6792	1.7926	1.9499	2.2151	2.5638
45	0.1847	0.3623	0.4301	0.4793	0.5201	0.5561	0.5889	1.5129	1.5868	1.6757	1.7883	1.9446	2.2078	2.5534
46	0.1851	0.3628	0.4305	0.4798	0.5206	0.5565	0.5892	1.5105	1.5839	1.6723	1.7843	1.9395	2.2008	2.5436
47	0.1854	0.3632	0.4310	0.4802	0.5210	0.5569	0.5896	1.5081	1.5812	1.6691	1.7804	1.9347	2.1941	2.5342
48	0.1857	0.3636	0.4314	0.4806	0.5214	0.5572	0.5899	1.5058	1.5786	1.6660	1.7768	1.9301	2.1878	2.5252
49	0.1859	0.3640	0.4318	0.4810	0.5217	0.5576	0.5903	1.5037	1.5761	1.6631	1.7732	1.9257	2.1817	2.5166
50	0.1862	0.3644	0.4322	0.4814	0.5221	0.5579	0.5906	1.5016	1.5736	1.6602	1.7698	1.9214	2.1758	2.5083

Tabla de Cuantiles de la Distribución F

Grados de libertad del numerador en extremo superior izquierdo, grados de libertad del denominador en margen izquierdo de cada fila. En el margen superior se lee $P(F \leq x)$ para los valores de x en el cuerpo de la tabla.

14	0.001	0.025	0.050	0.075	0.100	0.125	0.150	0.850	0.875	0.900	0.925	0.950	0.975	0.990
1	0.0583	0.1588	0.2174	0.2703	0.3224	0.3755	0.4311	26.9440	38.9573	61.0726	108.852	245.363	982.545	6143.00
2	0.0849	0.2059	0.2675	0.3190	0.3668	0.4130	0.4589	6.0820	7.4177	9.4200	12.7555	19.4243	39.4266	99.4260
3	0.1028	0.2358	0.2991	0.3503	0.3965	0.4401	0.4826	3.7842	4.3764	5.2047	6.4693	8.7149	14.2768	26.9238
4	0.1160	0.2569	0.3213	0.3723	0.4176	0.4598	0.5003	2.9871	3.3661	3.8776	4.6245	5.8733	8.6837	14.2486
5	0.1262	0.2730	0.3380	0.3889	0.4335	0.4746	0.5138	2.5887	2.8723	3.2468	3.7786	4.6358	6.4556	9.7700
6	0.1345	0.2856	0.3512	0.4018	0.4459	0.4863	0.5246	2.3504	2.5810	2.8809	3.2990	3.9559	5.2968	7.6050
7	0.1413	0.2959	0.3618	0.4123	0.4560	0.4958	0.5333	2.1918	2.3890	2.6426	2.9913	3.5292	4.5961	6.3590
8	0.1470	0.3044	0.3706	0.4210	0.4643	0.5036	0.5405	2.0785	2.2529	2.4752	2.7775	3.2374	4.1297	5.5588
9	0.1519	0.3116	0.3780	0.4282	0.4713	0.5102	0.5466	1.9936	2.1514	2.3510	2.6204	3.0255	3.7980	5.0052
10	0.1561	0.3178	0.3843	0.4345	0.4772	0.5158	0.5518	1.9275	2.0726	2.2553	2.5000	2.8647	3.5504	4.6008
11	0.1599	0.3231	0.3898	0.4398	0.4824	0.5207	0.5563	1.8745	2.0098	2.1792	2.4049	2.7386	3.3588	4.2933
12	0.1631	0.3279	0.3946	0.4446	0.4869	0.5250	0.5603	1.8311	1.9584	2.1173	2.3278	2.6371	3.2062	4.0517
13	0.1660	0.3320	0.3988	0.4487	0.4909	0.5287	0.5638	1.7948	1.9156	2.0658	2.2640	2.5536	3.0819	3.8573
14	0.1686	0.3357	0.4026	0.4524	0.4945	0.5321	0.5669	1.7641	1.8795	2.0224	2.2104	2.4837	2.9786	3.6976
15	0.1710	0.3391	0.4060	0.4557	0.4976	0.5351	0.5696	1.7377	1.8484	1.9853	2.1647	2.4244	2.8915	3.5639
16	0.1731	0.3421	0.4091	0.4587	0.5005	0.5378	0.5721	1.7148	1.8216	1.9532	2.1252	2.3733	2.8170	3.4506
17	0.1750	0.3448	0.4118	0.4614	0.5031	0.5402	0.5744	1.6948	1.7980	1.9252	2.0908	2.3290	2.7526	3.3533
18	0.1768	0.3473	0.4144	0.4639	0.5054	0.5424	0.5765	1.6770	1.7773	1.9004	2.0605	2.2900	2.6964	3.2689
19	0.1785	0.3496	0.4167	0.4661	0.5076	0.5445	0.5784	1.6612	1.7588	1.8785	2.0337	2.2556	2.6469	3.1949
20	0.1800	0.3517	0.4188	0.4682	0.5096	0.5463	0.5801	1.6471	1.7423	1.8588	2.0098	2.2250	2.6030	3.1296
21	0.1813	0.3536	0.4207	0.4701	0.5114	0.5480	0.5817	1.6343	1.7274	1.8412	1.9883	2.1975	2.5638	3.0715
22	0.1826	0.3554	0.4225	0.4719	0.5131	0.5496	0.5832	1.6228	1.7139	1.8252	1.9688	2.1727	2.5285	3.0195
23	0.1838	0.3570	0.4242	0.4735	0.5146	0.5511	0.5845	1.6122	1.7016	1.8107	1.9512	2.1502	2.4966	2.9727
24	0.1849	0.3586	0.4258	0.4750	0.5161	0.5525	0.5858	1.6026	1.6904	1.7974	1.9351	2.1298	2.4677	2.9303
25	0.1860	0.3600	0.4272	0.4764	0.5174	0.5537	0.5870	1.5938	1.6802	1.7853	1.9204	2.1111	2.4413	2.8917
26	0.1870	0.3614	0.4286	0.4777	0.5187	0.5549	0.5881	1.5856	1.6707	1.7741	1.9068	2.0939	2.4171	2.8566
27	0.1879	0.3626	0.4299	0.4790	0.5199	0.5561	0.5892	1.5781	1.6620	1.7638	1.8943	2.0781	2.3949	2.8243
28	0.1887	0.3638	0.4311	0.4801	0.5210	0.5571	0.5901	1.5712	1.6539	1.7542	1.8828	2.0635	2.3743	2.7946
29	0.1896	0.3649	0.4322	0.4812	0.5220	0.5581	0.5910	1.5647	1.6463	1.7454	1.8721	2.0500	2.3554	2.7672
30	0.1903	0.3660	0.4332	0.4823	0.5230	0.5590	0.5919	1.5586	1.6393	1.7371	1.8621	2.0374	2.3378	2.7418
31	0.1911	0.3670	0.4342	0.4832	0.5240	0.5599	0.5927	1.5530	1.6328	1.7294	1.8528	2.0257	2.3214	2.7182
32	0.1917	0.3679	0.4352	0.4842	0.5248	0.5607	0.5935	1.5477	1.6267	1.7222	1.8441	2.0147	2.3061	2.6963
33	0.1924	0.3688	0.4361	0.4850	0.5257	0.5615	0.5942	1.5428	1.6209	1.7154	1.8360	2.0045	2.2918	2.6758
34	0.1930	0.3697	0.4369	0.4859	0.5265	0.5623	0.5949	1.5381	1.6155	1.7091	1.8283	1.9949	2.2784	2.6566
35	0.1936	0.3705	0.4377	0.4866	0.5272	0.5630	0.5956	1.5337	1.6104	1.7031	1.8211	1.9858	2.2659	2.6387
36	0.1942	0.3712	0.4385	0.4874	0.5279	0.5636	0.5962	1.5296	1.6056	1.6974	1.8143	1.9773	2.2540	2.6218
37	0.1947	0.3719	0.4392	0.4881	0.5286	0.5643	0.5968	1.5257	1.6011	1.6921	1.8079	1.9692	2.2429	2.6059
38	0.1952	0.3726	0.4399	0.4888	0.5293	0.5649	0.5974	1.5220	1.5968	1.6871	1.8019	1.9616	2.2324	2.5909
39	0.1957	0.3733	0.4406	0.4894	0.5299	0.5655	0.5979	1.5184	1.5927	1.6823	1.7961	1.9545	2.2224	2.5768
40	0.1962	0.3739	0.4412	0.4900	0.5305	0.5660	0.5984	1.5151	1.5889	1.6778	1.7907	1.9476	2.2130	2.5634
41	0.1966	0.3745	0.4418	0.4906	0.5310	0.5665	0.5989	1.5119	1.5852	1.6735	1.7855	1.9412	2.2040	2.5507
42	0.1970	0.3751	0.4424	0.4912	0.5316	0.5671	0.5994	1.5089	1.5817	1.6694	1.7806	1.9350	2.1955	2.5387
43	0.1974	0.3757	0.4430	0.4917	0.5321	0.5675	0.5998	1.5060	1.5784	1.6655	1.7759	1.9292	2.1875	2.5273
44	0.1978	0.3762	0.4435	0.4923	0.5326	0.5680	0.6003	1.5033	1.5752	1.6618	1.7715	1.9236	2.1798	2.5164
45	0.1982	0.3767	0.4440	0.4928	0.5331	0.5685	0.6007	1.5007	1.5722	1.6582	1.7672	1.9182	2.1725	2.5061
46	0.1986	0.3772	0.4445	0.4932	0.5335	0.5689	0.6011	1.4982	1.5693	1.6548	1.7631	1.9132	2.1655	2.4962
47	0.1989	0.3777	0.4450	0.4937	0.5339	0.5693	0.6015	1.4958	1.5665	1.6516	1.7592	1.9083	2.1588	2.4868
48	0.1993	0.3782	0.4455	0.4941	0.5344	0.5697	0.6019	1.4935	1.5639	1.6485	1.7555	1.9037	2.1524	2.4777
49	0.1996	0.3786	0.4459	0.4946	0.5348	0.5701	0.6022	1.4913	1.5613	1.6455	1.7519	1.8992	2.1463	2.4691
50	0.1999	0.3790	0.4463	0.4950	0.5352	0.5705	0.6026	1.4891	1.5589	1.6426	1.7485	1.8949	2.1404	2.4609

Tabla de Cuantiles de la Distribución F

Grados de libertad del numerador en extremo superior izquierdo, grados de libertad del denominador en margen izquierdo de cada fila. En el margen superior se lee $P(F \leq x)$ para los valores de x en el cuerpo de la tabla.

15	0.001	0.025	0.050	0.075	0.100	0.125	0.150	0.850	0.875	0.900	0.925	0.950	0.975	0.990
1	0.0603	0.1613	0.2201	0.2732	0.3254	0.3787	0.4344	27.0106	39.0526	61.2204	109.113	245.949	984.874	6156.97
2	0.0882	0.2099	0.2716	0.3232	0.3710	0.4173	0.4633	6.0867	7.4224	9.4247	12.7603	19.4291	39.4311	99.4332
3	0.1071	0.2408	0.3042	0.3555	0.4016	0.4453	0.4877	3.7828	4.3738	5.2003	6.4622	8.7028	14.2527	26.8719
4	0.1212	0.2629	0.3273	0.3783	0.4235	0.4656	0.5060	2.9835	3.3610	3.8704	4.6142	5.8578	8.6566	14.1981
5	0.1321	0.2796	0.3447	0.3955	0.4399	0.4810	0.5201	2.5838	2.8658	3.2380	3.7667	4.6188	6.4277	9.7223
6	0.1410	0.2929	0.3584	0.4089	0.4529	0.4932	0.5312	2.3446	2.5736	2.8712	3.2861	3.9381	5.2686	7.5590
7	0.1483	0.3036	0.3695	0.4199	0.4634	0.5030	0.5403	2.1854	2.3809	2.6322	2.9777	3.5107	4.5678	6.3144
8	0.1545	0.3126	0.3787	0.4289	0.4720	0.5112	0.5478	2.0717	2.2443	2.4642	2.7634	3.2184	4.1012	5.5152
9	0.1598	0.3202	0.3865	0.4365	0.4793	0.5180	0.5542	1.9863	2.1423	2.3396	2.6058	3.0061	3.7693	4.9621
10	0.1644	0.3268	0.3931	0.4431	0.4856	0.5239	0.5596	1.9198	2.0632	2.2435	2.4851	2.8450	3.5217	4.5582
11	0.1685	0.3325	0.3989	0.4487	0.4910	0.5291	0.5644	1.8666	2.0000	2.1671	2.3896	2.7186	3.3299	4.2509
12	0.1721	0.3375	0.4040	0.4537	0.4958	0.5335	0.5685	1.8229	1.9483	2.1049	2.3122	2.6169	3.1772	4.0096
13	0.1752	0.3419	0.4085	0.4581	0.5000	0.5375	0.5722	1.7864	1.9053	2.0532	2.2482	2.5331	3.0527	3.8154
14	0.1781	0.3458	0.4125	0.4620	0.5037	0.5410	0.5755	1.7555	1.8689	2.0095	2.1943	2.4630	2.9493	3.6557
15	0.1807	0.3494	0.4161	0.4655	0.5070	0.5441	0.5784	1.7289	1.8377	1.9722	2.1484	2.4034	2.8621	3.5222
16	0.1830	0.3526	0.4193	0.4686	0.5101	0.5470	0.5810	1.7059	1.8107	1.9399	2.1087	2.3522	2.7875	3.4090
17	0.1851	0.3555	0.4222	0.4715	0.5128	0.5496	0.5834	1.6856	1.7870	1.9117	2.0742	2.3077	2.7230	3.3117
18	0.1871	0.3582	0.4249	0.4741	0.5153	0.5519	0.5856	1.6678	1.7661	1.8868	2.0438	2.2686	2.6667	3.2273
19	0.1889	0.3606	0.4274	0.4765	0.5176	0.5541	0.5876	1.6518	1.7475	1.8647	2.0168	2.2341	2.6171	3.1533
20	0.1905	0.3629	0.4296	0.4787	0.5197	0.5560	0.5894	1.6376	1.7308	1.8449	1.9927	2.2033	2.5731	3.0880
21	0.1921	0.3649	0.4317	0.4807	0.5216	0.5579	0.5911	1.6247	1.7158	1.8271	1.9710	2.1757	2.5338	3.0300
22	0.1935	0.3668	0.4336	0.4826	0.5234	0.5595	0.5927	1.6130	1.7022	1.8111	1.9515	2.1508	2.4984	2.9779
23	0.1948	0.3686	0.4354	0.4843	0.5250	0.5611	0.5941	1.6024	1.6898	1.7964	1.9337	2.1282	2.4665	2.9311
24	0.1961	0.3703	0.4371	0.4859	0.5266	0.5626	0.5955	1.5927	1.6785	1.7831	1.9176	2.1077	2.4374	2.8887
25	0.1972	0.3718	0.4386	0.4874	0.5280	0.5639	0.5967	1.5838	1.6682	1.7708	1.9027	2.0889	2.4110	2.8502
26	0.1983	0.3733	0.4401	0.4888	0.5294	0.5652	0.5979	1.5756	1.6586	1.7596	1.8891	2.0716	2.3867	2.8150
27	0.1993	0.3746	0.4415	0.4901	0.5306	0.5664	0.5990	1.5680	1.6498	1.7492	1.8765	2.0558	2.3644	2.7827
28	0.2003	0.3759	0.4427	0.4914	0.5318	0.5675	0.6001	1.5610	1.6416	1.7395	1.8649	2.0411	2.3438	2.7530
29	0.2012	0.3771	0.4439	0.4926	0.5329	0.5685	0.6010	1.5544	1.6341	1.7306	1.8541	2.0275	2.3248	2.7256
30	0.2020	0.3783	0.4451	0.4937	0.5340	0.5695	0.6020	1.5483	1.6270	1.7223	1.8441	2.0148	2.3072	2.7002
31	0.2028	0.3793	0.4462	0.4947	0.5350	0.5704	0.6028	1.5426	1.6204	1.7145	1.8347	2.0030	2.2907	2.6766
32	0.2036	0.3803	0.4472	0.4957	0.5359	0.5713	0.6036	1.5373	1.6142	1.7072	1.8259	1.9920	2.2754	2.6546
33	0.2043	0.3813	0.4481	0.4966	0.5368	0.5722	0.6044	1.5323	1.6084	1.7004	1.8177	1.9817	2.2610	2.6341
34	0.2050	0.3822	0.4490	0.4975	0.5376	0.5729	0.6052	1.5276	1.6029	1.6940	1.8100	1.9720	2.2476	2.6150
35	0.2057	0.3831	0.4499	0.4983	0.5384	0.5737	0.6059	1.5231	1.5978	1.6880	1.8028	1.9629	2.2350	2.5970
36	0.2063	0.3839	0.4507	0.4991	0.5392	0.5744	0.6065	1.5189	1.5929	1.6823	1.7959	1.9543	2.2231	2.5801
37	0.2069	0.3847	0.4515	0.4999	0.5399	0.5751	0.6072	1.5150	1.5884	1.6769	1.7895	1.9462	2.2119	2.5642
38	0.2075	0.3854	0.4523	0.5006	0.5406	0.5758	0.6078	1.5112	1.5840	1.6718	1.7834	1.9386	2.2014	2.5492
39	0.2080	0.3862	0.4530	0.5013	0.5413	0.5764	0.6083	1.5077	1.5799	1.6670	1.7776	1.9313	2.1914	2.5350
40	0.2085	0.3868	0.4537	0.5020	0.5419	0.5770	0.6089	1.5043	1.5760	1.6624	1.7721	1.9245	2.1819	2.5216
41	0.2090	0.3875	0.4543	0.5026	0.5425	0.5775	0.6094	1.5011	1.5723	1.6581	1.7669	1.9179	2.1729	2.5089
42	0.2095	0.3881	0.4549	0.5032	0.5431	0.5781	0.6099	1.4980	1.5688	1.6539	1.7619	1.9118	2.1644	2.4969
43	0.2100	0.3887	0.4555	0.5038	0.5436	0.5786	0.6104	1.4951	1.5654	1.6500	1.7572	1.9059	2.1563	2.4854
44	0.2104	0.3893	0.4561	0.5043	0.5442	0.5791	0.6109	1.4923	1.5622	1.6462	1.7527	1.9002	2.1486	2.4746
45	0.2108	0.3899	0.4567	0.5049	0.5447	0.5796	0.6113	1.4897	1.5591	1.6426	1.7484	1.8949	2.1412	2.4642
46	0.2112	0.3904	0.4572	0.5054	0.5452	0.5800	0.6117	1.4872	1.5562	1.6392	1.7443	1.8898	2.1342	2.4543
47	0.2116	0.3909	0.4577	0.5059	0.5456	0.5805	0.6122	1.4847	1.5534	1.6359	1.7404	1.8849	2.1275	2.4449
48	0.2120	0.3914	0.4582	0.5064	0.5461	0.5809	0.6126	1.4824	1.5507	1.6328	1.7366	1.8802	2.1210	2.4359
49	0.2123	0.3919	0.4587	0.5068	0.5465	0.5813	0.6129	1.4802	1.5482	1.6298	1.7330	1.8757	2.1149	2.4272
50	0.2127	0.3923	0.4591	0.5073	0.5469	0.5817	0.6133	1.4780	1.5457	1.6269	1.7296	1.8714	2.1090	2.4190

Tabla de Cuantiles de la Distribución F

Grados de libertad del numerador en extremo superior izquierdo, grados de libertad del denominador en margen izquierdo de cada fila. En el margen superior se lee P(F ≤ x) para los valores de x en el cuerpo de la tabla.

16	0.001	0.025	0.050	0.075	0.100	0.125	0.150	0.850	0.875	0.900	0.925	0.950	0.975	0.990
1	0.0620	0.1635	0.2225	0.2757	0.3281	0.3815	0.4373	27.0688	39.1360	61.3500	109.343	246.466	986.911	6170.01
2	0.0911	0.2134	0.2752	0.3269	0.3748	0.4211	0.4671	6.0908	7.4265	9.4288	12.7644	19.4332	39.4357	99.4369
3	0.1110	0.2453	0.3087	0.3600	0.4062	0.4498	0.4922	3.7816	4.3714	5.1964	6.4559	8.6923	14.2315	26.8265
4	0.1259	0.2681	0.3326	0.3835	0.4287	0.4707	0.5111	2.9803	3.3564	3.8639	4.6050	5.8441	8.6326	14.1540
5	0.1375	0.2855	0.3506	0.4013	0.4457	0.4866	0.5256	2.5795	2.8600	3.2303	3.7562	4.6038	6.4032	9.6802
6	0.1470	0.2993	0.3648	0.4153	0.4591	0.4992	0.5371	2.3396	2.5670	2.8626	3.2747	3.9223	5.2439	7.5186
7	0.1548	0.3106	0.3763	0.4266	0.4699	0.5094	0.5465	2.1798	2.3737	2.6230	2.9657	3.4944	4.5428	6.2751
8	0.1614	0.3200	0.3859	0.4360	0.4789	0.5179	0.5543	2.0656	2.2366	2.4545	2.7509	3.2016	4.0761	5.4765
9	0.1671	0.3280	0.3941	0.4439	0.4865	0.5250	0.5610	1.9799	2.1342	2.3295	2.5929	2.9890	3.7441	4.9240
10	0.1721	0.3349	0.4010	0.4508	0.4931	0.5312	0.5666	1.9131	2.0548	2.2330	2.4718	2.8276	3.4963	4.5204
11	0.1764	0.3409	0.4071	0.4567	0.4987	0.5365	0.5716	1.8595	1.9913	2.1563	2.3760	2.7009	3.3044	4.2135
12	0.1803	0.3461	0.4124	0.4619	0.5037	0.5412	0.5759	1.8156	1.9394	2.0938	2.2984	2.5989	3.1515	3.9724
13	0.1837	0.3508	0.4171	0.4664	0.5081	0.5453	0.5797	1.7789	1.8962	2.0419	2.2341	2.5149	3.0269	3.7783
14	0.1868	0.3550	0.4214	0.4705	0.5120	0.5490	0.5831	1.7478	1.8596	1.9981	2.1801	2.4446	2.9234	3.6187
15	0.1896	0.3587	0.4251	0.4742	0.5155	0.5523	0.5862	1.7211	1.8282	1.9605	2.1339	2.3849	2.8360	3.4852
16	0.1921	0.3621	0.4285	0.4775	0.5187	0.5553	0.5890	1.6979	1.8010	1.9281	2.0941	2.3335	2.7614	3.3721
17	0.1944	0.3652	0.4316	0.4805	0.5215	0.5580	0.5915	1.6775	1.7771	1.8997	2.0594	2.2888	2.6968	3.2748
18	0.1966	0.3681	0.4345	0.4833	0.5241	0.5604	0.5938	1.6595	1.7561	1.8747	2.0288	2.2496	2.6403	3.1905
19	0.1985	0.3706	0.4371	0.4858	0.5265	0.5627	0.5959	1.6434	1.7373	1.8524	2.0017	2.2149	2.5907	3.1165
20	0.2003	0.3730	0.4395	0.4881	0.5287	0.5648	0.5978	1.6291	1.7206	1.8325	1.9775	2.1840	2.5465	3.0512
21	0.2020	0.3752	0.4417	0.4902	0.5308	0.5667	0.5996	1.6161	1.7054	1.8146	1.9557	2.1563	2.5071	2.9931
22	0.2036	0.3773	0.4437	0.4922	0.5327	0.5684	0.6012	1.6043	1.6918	1.7984	1.9360	2.1313	2.4717	2.9411
23	0.2050	0.3792	0.4456	0.4941	0.5344	0.5701	0.6027	1.5936	1.6793	1.7837	1.9182	2.1086	2.4396	2.8942
24	0.2063	0.3809	0.4473	0.4958	0.5360	0.5716	0.6042	1.5838	1.6679	1.7703	1.9019	2.0880	2.4105	2.8519
25	0.2076	0.3826	0.4490	0.4974	0.5375	0.5730	0.6055	1.5748	1.6575	1.7579	1.8870	2.0691	2.3840	2.8133
26	0.2088	0.3841	0.4505	0.4988	0.5390	0.5744	0.6067	1.5666	1.6478	1.7466	1.8733	2.0518	2.3597	2.7781
27	0.2099	0.3856	0.4520	0.5002	0.5403	0.5756	0.6079	1.5589	1.6389	1.7361	1.8606	2.0358	2.3373	2.7458
28	0.2110	0.3869	0.4533	0.5016	0.5415	0.5768	0.6090	1.5518	1.6307	1.7264	1.8489	2.0210	2.3167	2.7160
29	0.2120	0.3882	0.4546	0.5028	0.5427	0.5779	0.6100	1.5452	1.6230	1.7174	1.8381	2.0073	2.2976	2.6886
30	0.2129	0.3894	0.4558	0.5040	0.5438	0.5790	0.6110	1.5390	1.6159	1.7090	1.8280	1.9946	2.2799	2.6632
31	0.2138	0.3906	0.4570	0.5051	0.5449	0.5799	0.6119	1.5333	1.6092	1.7012	1.8185	1.9828	2.2634	2.6396
32	0.2146	0.3917	0.4580	0.5061	0.5459	0.5809	0.6128	1.5279	1.6030	1.6938	1.8097	1.9717	2.2480	2.6176
33	0.2154	0.3927	0.4591	0.5071	0.5468	0.5818	0.6136	1.5228	1.5971	1.6869	1.8014	1.9613	2.2336	2.5971
34	0.2162	0.3937	0.4600	0.5080	0.5477	0.5826	0.6144	1.5181	1.5916	1.6805	1.7937	1.9516	2.2201	2.5779
35	0.2169	0.3946	0.4610	0.5089	0.5486	0.5834	0.6151	1.5136	1.5864	1.6744	1.7863	1.9424	2.2075	2.5599
36	0.2176	0.3955	0.4618	0.5098	0.5494	0.5841	0.6158	1.5093	1.5816	1.6687	1.7795	1.9338	2.1956	2.5430
37	0.2183	0.3963	0.4627	0.5106	0.5501	0.5849	0.6165	1.5053	1.5769	1.6632	1.7729	1.9256	2.1843	2.5270
38	0.2189	0.3971	0.4635	0.5113	0.5509	0.5856	0.6171	1.5016	1.5726	1.6581	1.7668	1.9179	2.1737	2.5120
39	0.2195	0.3979	0.4642	0.5121	0.5516	0.5862	0.6177	1.4980	1.5684	1.6532	1.7610	1.9107	2.1637	2.4978
40	0.2200	0.3986	0.4650	0.5128	0.5522	0.5868	0.6183	1.4946	1.5645	1.6486	1.7554	1.9038	2.1542	2.4844
41	0.2206	0.3993	0.4657	0.5135	0.5529	0.5874	0.6189	1.4913	1.5607	1.6442	1.7502	1.8972	2.1452	2.4717
42	0.2211	0.4000	0.4663	0.5141	0.5535	0.5880	0.6194	1.4882	1.5572	1.6401	1.7452	1.8910	2.1366	2.4596
43	0.2216	0.4007	0.4670	0.5147	0.5541	0.5886	0.6199	1.4853	1.5538	1.6361	1.7404	1.8850	2.1285	2.4482
44	0.2221	0.4013	0.4676	0.5153	0.5546	0.5891	0.6204	1.4825	1.5505	1.6323	1.7359	1.8794	2.1207	2.4373
45	0.2226	0.4019	0.4682	0.5159	0.5552	0.5896	0.6209	1.4798	1.5474	1.6287	1.7316	1.8740	2.1133	2.4269
46	0.2230	0.4025	0.4687	0.5164	0.5557	0.5901	0.6213	1.4772	1.5445	1.6252	1.7274	1.8688	2.1063	2.4170
47	0.2235	0.4030	0.4693	0.5170	0.5562	0.5906	0.6218	1.4748	1.5416	1.6219	1.7235	1.8639	2.0995	2.4075
48	0.2239	0.4035	0.4698	0.5175	0.5567	0.5910	0.6222	1.4724	1.5389	1.6187	1.7197	1.8592	2.0931	2.3985
49	0.2243	0.4040	0.4703	0.5179	0.5571	0.5915	0.6226	1.4702	1.5363	1.6157	1.7160	1.8546	2.0869	2.3899
50	0.2247	0.4045	0.4708	0.5184	0.5576	0.5919	0.6230	1.4680	1.5338	1.6128	1.7125	1.8503	2.0810	2.3816

Tabla de Cuantiles de la Distribución F

Grados de libertad del numerador en extremo superior izquierdo, grados de libertad del denominador en margen izquierdo de cada fila. En el margen superior se lee P(F ≤ x) para los valores de x en el cuerpo de la tabla.

17	0.001	0.025	0.050	0.075	0.100	0.125	0.150	0.850	0.875	0.900	0.925	0.950	0.975	0.990
1	0.0636	0.1655	0.2247	0.2780	0.3304	0.3840	0.4399	27.1204	39.2097	61.4646	109.545	246.917	988.715	6181.19
2	0.0938	0.2165	0.2784	0.3302	0.3781	0.4245	0.4705	6.0945	7.4302	9.4325	12.7681	19.4370	39.4393	99.4405
3	0.1146	0.2493	0.3128	0.3641	0.4103	0.4539	0.4962	3.7805	4.3694	5.1929	6.4503	8.6829	14.2127	26.7864
4	0.1302	0.2729	0.3373	0.3882	0.4333	0.4753	0.5156	2.9774	3.3523	3.8582	4.5969	5.8320	8.6113	14.1144
5	0.1424	0.2909	0.3559	0.4065	0.4508	0.4917	0.5305	2.5756	2.8549	3.2234	3.7469	4.5904	6.3814	9.6429
6	0.1524	0.3052	0.3706	0.4209	0.4646	0.5046	0.5424	2.3350	2.5611	2.8550	3.2646	3.9083	5.2218	7.4826
7	0.1607	0.3169	0.3825	0.4326	0.4758	0.5151	0.5521	2.1747	2.3673	2.6148	2.9550	3.4799	4.5206	6.2400
8	0.1677	0.3267	0.3925	0.4424	0.4851	0.5239	0.5602	2.0601	2.2297	2.4458	2.7397	3.1867	4.0538	5.4423
9	0.1738	0.3350	0.4009	0.4506	0.4930	0.5313	0.5670	1.9741	2.1270	2.3205	2.5814	2.9737	3.7216	4.8902
10	0.1791	0.3422	0.4082	0.4577	0.4998	0.5377	0.5729	1.9070	2.0472	2.2237	2.4600	2.8120	3.4736	4.4869
11	0.1837	0.3485	0.4145	0.4638	0.5056	0.5432	0.5780	1.8532	1.9835	2.1467	2.3639	2.6851	3.2816	4.1802
12	0.1878	0.3540	0.4201	0.4692	0.5108	0.5480	0.5825	1.8091	1.9314	2.0839	2.2860	2.5828	3.1286	3.9392
13	0.1915	0.3589	0.4250	0.4740	0.5154	0.5523	0.5865	1.7722	1.8879	2.0318	2.2215	2.4987	3.0039	3.7452
14	0.1948	0.3633	0.4294	0.4783	0.5194	0.5562	0.5901	1.7409	1.8512	1.9878	2.1673	2.4282	2.9003	3.5857
15	0.1978	0.3672	0.4333	0.4821	0.5231	0.5596	0.5932	1.7140	1.8196	1.9501	2.1210	2.3683	2.8128	3.4523
16	0.2006	0.3708	0.4369	0.4856	0.5264	0.5627	0.5961	1.6906	1.7922	1.9175	2.0810	2.3167	2.7380	3.3392
17	0.2031	0.3741	0.4402	0.4887	0.5294	0.5655	0.5987	1.6702	1.7683	1.8889	2.0461	2.2719	2.6733	3.2419
18	0.2054	0.3770	0.4431	0.4916	0.5321	0.5681	0.6011	1.6520	1.7471	1.8638	2.0154	2.2325	2.6168	3.1575
19	0.2075	0.3798	0.4459	0.4942	0.5346	0.5705	0.6033	1.6359	1.7282	1.8414	1.9881	2.1977	2.5670	3.0836
20	0.2094	0.3823	0.4484	0.4967	0.5370	0.5726	0.6053	1.6214	1.7113	1.8214	1.9638	2.1667	2.5228	3.0183
21	0.2112	0.3846	0.4507	0.4989	0.5391	0.5746	0.6072	1.6083	1.6961	1.8034	1.9419	2.1389	2.4833	2.9602
22	0.2129	0.3868	0.4528	0.5010	0.5411	0.5765	0.6089	1.5965	1.6823	1.7871	1.9222	2.1138	2.4478	2.9082
23	0.2145	0.3888	0.4548	0.5029	0.5429	0.5782	0.6105	1.5857	1.6698	1.7723	1.9042	2.0910	2.4156	2.8613
24	0.2159	0.3907	0.4567	0.5047	0.5446	0.5798	0.6120	1.5758	1.6583	1.7587	1.8879	2.0703	2.3865	2.8189
25	0.2173	0.3924	0.4584	0.5064	0.5462	0.5813	0.6134	1.5668	1.6478	1.7463	1.8729	2.0513	2.3599	2.7803
26	0.2186	0.3940	0.4600	0.5079	0.5477	0.5827	0.6147	1.5584	1.6381	1.7349	1.8591	2.0339	2.3355	2.7451
27	0.2198	0.3956	0.4616	0.5094	0.5491	0.5840	0.6159	1.5507	1.6291	1.7243	1.8463	2.0179	2.3131	2.7127
28	0.2210	0.3970	0.4630	0.5108	0.5504	0.5853	0.6171	1.5435	1.6208	1.7146	1.8346	2.0030	2.2924	2.6830
29	0.2220	0.3984	0.4643	0.5121	0.5516	0.5864	0.6181	1.5368	1.6131	1.7055	1.8236	1.9893	2.2732	2.6555
30	0.2230	0.3997	0.4656	0.5133	0.5528	0.5875	0.6192	1.5306	1.6059	1.6970	1.8134	1.9765	2.2554	2.6301
31	0.2240	0.4009	0.4668	0.5145	0.5539	0.5885	0.6201	1.5248	1.5992	1.6891	1.8040	1.9646	2.2389	2.6064
32	0.2249	0.4020	0.4680	0.5156	0.5549	0.5895	0.6210	1.5194	1.5929	1.6818	1.7951	1.9534	2.2235	2.5844
33	0.2258	0.4031	0.4690	0.5166	0.5559	0.5905	0.6219	1.5143	1.5870	1.6748	1.7867	1.9430	2.2090	2.5639
34	0.2266	0.4042	0.4701	0.5176	0.5569	0.5913	0.6227	1.5095	1.5814	1.6683	1.7789	1.9332	2.1955	2.5447
35	0.2274	0.4052	0.4710	0.5185	0.5578	0.5922	0.6235	1.5050	1.5762	1.6622	1.7716	1.9240	2.1828	2.5266
36	0.2281	0.4061	0.4720	0.5194	0.5586	0.5930	0.6242	1.5007	1.5713	1.6564	1.7646	1.9153	2.1708	2.5097
37	0.2289	0.4070	0.4729	0.5203	0.5594	0.5937	0.6249	1.4966	1.5666	1.6509	1.7581	1.9071	2.1595	2.4938
38	0.2295	0.4079	0.4737	0.5211	0.5602	0.5945	0.6256	1.4928	1.5622	1.6457	1.7519	1.8994	2.1489	2.4787
39	0.2302	0.4087	0.4745	0.5219	0.5609	0.5952	0.6263	1.4892	1.5580	1.6408	1.7460	1.8921	2.1388	2.4645
40	0.2308	0.4095	0.4753	0.5226	0.5616	0.5958	0.6269	1.4857	1.5540	1.6362	1.7404	1.8851	2.1293	2.4511
41	0.2314	0.4102	0.4760	0.5233	0.5623	0.5964	0.6275	1.4825	1.5502	1.6318	1.7351	1.8785	2.1202	2.4384
42	0.2320	0.4109	0.4767	0.5240	0.5630	0.5971	0.6280	1.4794	1.5466	1.6276	1.7301	1.8722	2.1116	2.4263
43	0.2325	0.4116	0.4774	0.5247	0.5636	0.5976	0.6286	1.4764	1.5432	1.6235	1.7253	1.8663	2.1035	2.4148
44	0.2331	0.4123	0.4781	0.5253	0.5642	0.5982	0.6291	1.4735	1.5399	1.6197	1.7207	1.8606	2.0957	2.4039
45	0.2336	0.4129	0.4787	0.5259	0.5647	0.5987	0.6296	1.4708	1.5368	1.6161	1.7164	1.8551	2.0882	2.3935
46	0.2341	0.4135	0.4793	0.5265	0.5653	0.5992	0.6301	1.4682	1.5338	1.6126	1.7122	1.8500	2.0812	2.3835
47	0.2346	0.4141	0.4799	0.5270	0.5658	0.5997	0.6305	1.4658	1.5310	1.6092	1.7082	1.8450	2.0744	2.3741
48	0.2350	0.4147	0.4804	0.5276	0.5663	0.6002	0.6310	1.4634	1.5282	1.6060	1.7044	1.8402	2.0679	2.3650
49	0.2354	0.4152	0.4810	0.5281	0.5668	0.6007	0.6314	1.4611	1.5256	1.6030	1.7007	1.8357	2.0617	2.3564
50	0.2359	0.4158	0.4815	0.5286	0.5673	0.6011	0.6318	1.4589	1.5231	1.6000	1.6972	1.8313	2.0558	2.3481

Tabla de Cuantiles de la Distribución F

Grados de libertad del numerador en extremo superior izquierdo, grados de libertad del denominador en margen izquierdo de cada fila. En el margen superior se lee P(F ≤ x) para los valores de x en el cuerpo de la tabla.

18	0.001	0.025	0.050	0.075	0.100	0.125	0.150	0.850	0.875	0.900	0.925	0.950	0.975	0.990
1	0.0650	0.1673	0.2266	0.2800	0.3326	0.3862	0.4423	27.1664	39.2754	61.5664	109.725	247.324	990.345	6191.43
2	0.0963	0.2193	0.2813	0.3332	0.3811	0.4275	0.4735	6.0977	7.4335	9.4358	12.7714	19.4402	39.4421	99.4442
3	0.1178	0.2529	0.3165	0.3678	0.4139	0.4575	0.4998	3.7795	4.3675	5.1898	6.4452	8.6745	14.1961	26.7510
4	0.1341	0.2771	0.3416	0.3925	0.4375	0.4794	0.5196	2.9748	3.3487	3.8531	4.5897	5.8211	8.5923	14.0794
5	0.1469	0.2957	0.3606	0.4112	0.4554	0.4962	0.5349	2.5722	2.8502	3.2172	3.7385	4.5785	6.3619	9.6095
6	0.1574	0.3105	0.3758	0.4260	0.4696	0.5095	0.5471	2.3309	2.5558	2.8481	3.2556	3.8957	5.2021	7.4506
7	0.1661	0.3226	0.3881	0.4381	0.4811	0.5203	0.5571	2.1702	2.3615	2.6074	2.9455	3.4669	4.5008	6.2089
8	0.1735	0.3327	0.3984	0.4481	0.4907	0.5293	0.5654	2.0552	2.2236	2.4380	2.7297	3.1733	4.0338	5.4116
9	0.1799	0.3414	0.4071	0.4566	0.4988	0.5369	0.5725	1.9689	2.1205	2.3123	2.5710	2.9600	3.7015	4.8599
10	0.1855	0.3489	0.4146	0.4639	0.5058	0.5435	0.5786	1.9015	2.0405	2.2153	2.4493	2.7980	3.4534	4.4569
11	0.1905	0.3554	0.4212	0.4703	0.5119	0.5492	0.5838	1.8475	1.9765	2.1380	2.3530	2.6709	3.2612	4.1503
12	0.1948	0.3612	0.4270	0.4759	0.5172	0.5542	0.5885	1.8032	1.9242	2.0750	2.2749	2.5684	3.1081	3.9095
13	0.1988	0.3663	0.4321	0.4809	0.5220	0.5587	0.5926	1.7661	1.8805	2.0227	2.2102	2.4841	2.9832	3.7156
14	0.2023	0.3709	0.4367	0.4853	0.5262	0.5627	0.5963	1.7347	1.8436	1.9785	2.1558	2.4134	2.8795	3.5561
15	0.2055	0.3750	0.4408	0.4893	0.5300	0.5662	0.5996	1.7076	1.8119	1.9407	2.1093	2.3533	2.7919	3.4228
16	0.2084	0.3787	0.4445	0.4929	0.5334	0.5695	0.6026	1.6841	1.7844	1.9079	2.0692	2.3016	2.7170	3.3096
17	0.2111	0.3821	0.4479	0.4962	0.5365	0.5724	0.6053	1.6635	1.7602	1.8792	2.0341	2.2567	2.6522	3.2124
18	0.2135	0.3853	0.4510	0.4992	0.5394	0.5751	0.6078	1.6453	1.7389	1.8539	2.0033	2.2172	2.5956	3.1280
19	0.2158	0.3881	0.4539	0.5019	0.5420	0.5775	0.6101	1.6290	1.7200	1.8314	1.9759	2.1823	2.5457	3.0541
20	0.2179	0.3908	0.4565	0.5045	0.5444	0.5798	0.6122	1.6144	1.7030	1.8113	1.9515	2.1511	2.5014	2.9887
21	0.2198	0.3932	0.4589	0.5068	0.5466	0.5819	0.6141	1.6013	1.6877	1.7932	1.9295	2.1232	2.4618	2.9306
22	0.2216	0.3955	0.4612	0.5090	0.5487	0.5838	0.6159	1.5894	1.6738	1.7768	1.9097	2.0980	2.4262	2.8786
23	0.2233	0.3976	0.4632	0.5110	0.5506	0.5856	0.6176	1.5785	1.6612	1.7619	1.8916	2.0751	2.3940	2.8317
24	0.2249	0.3996	0.4652	0.5128	0.5524	0.5873	0.6191	1.5686	1.6496	1.7483	1.8752	2.0543	2.3648	2.7892
25	0.2263	0.4014	0.4670	0.5146	0.5540	0.5888	0.6206	1.5594	1.6390	1.7358	1.8601	2.0353	2.3381	2.7506
26	0.2277	0.4031	0.4687	0.5162	0.5556	0.5903	0.6219	1.5510	1.6293	1.7243	1.8462	2.0178	2.3137	2.7154
27	0.2290	0.4048	0.4703	0.5178	0.5571	0.5916	0.6232	1.5432	1.6202	1.7137	1.8334	2.0017	2.2912	2.6830
28	0.2302	0.4063	0.4718	0.5192	0.5584	0.5929	0.6244	1.5360	1.6119	1.7039	1.8216	1.9868	2.2704	2.6532
29	0.2314	0.4077	0.4732	0.5206	0.5597	0.5942	0.6255	1.5293	1.6041	1.6947	1.8106	1.9730	2.2512	2.6257
30	0.2325	0.4091	0.4746	0.5219	0.5610	0.5953	0.6266	1.5230	1.5968	1.6862	1.8003	1.9601	2.2334	2.6002
31	0.2335	0.4104	0.4758	0.5231	0.5621	0.5964	0.6276	1.5172	1.5901	1.6783	1.7908	1.9481	2.2168	2.5766
32	0.2345	0.4116	0.4770	0.5242	0.5632	0.5974	0.6285	1.5117	1.5837	1.6708	1.7818	1.9369	2.2013	2.5546
33	0.2355	0.4127	0.4782	0.5253	0.5642	0.5984	0.6294	1.5065	1.5778	1.6638	1.7734	1.9264	2.1868	2.5340
34	0.2364	0.4138	0.4793	0.5264	0.5652	0.5993	0.6303	1.5017	1.5722	1.6573	1.7656	1.9166	2.1732	2.5148
35	0.2372	0.4149	0.4803	0.5274	0.5662	0.6002	0.6311	1.4971	1.5669	1.6511	1.7582	1.9073	2.1605	2.4967
36	0.2380	0.4159	0.4813	0.5283	0.5670	0.6010	0.6319	1.4928	1.5619	1.6453	1.7512	1.8986	2.1485	2.4797
37	0.2388	0.4168	0.4822	0.5292	0.5679	0.6018	0.6326	1.4887	1.5572	1.6397	1.7446	1.8904	2.1372	2.4638
38	0.2395	0.4177	0.4831	0.5300	0.5687	0.6026	0.6333	1.4849	1.5528	1.6345	1.7383	1.8826	2.1265	2.4487
39	0.2402	0.4186	0.4839	0.5309	0.5695	0.6033	0.6340	1.4812	1.5485	1.6296	1.7324	1.8752	2.1164	2.4345
40	0.2409	0.4194	0.4848	0.5316	0.5702	0.6040	0.6347	1.4777	1.5445	1.6249	1.7268	1.8682	2.1068	2.4210
41	0.2416	0.4202	0.4855	0.5324	0.5709	0.6047	0.6353	1.4744	1.5407	1.6204	1.7215	1.8616	2.0977	2.4083
42	0.2422	0.4210	0.4863	0.5331	0.5716	0.6053	0.6359	1.4713	1.5371	1.6162	1.7164	1.8553	2.0891	2.3961
43	0.2428	0.4217	0.4870	0.5338	0.5723	0.6059	0.6364	1.4683	1.5336	1.6121	1.7116	1.8493	2.0809	2.3847
44	0.2434	0.4224	0.4877	0.5345	0.5729	0.6065	0.6370	1.4654	1.5303	1.6083	1.7070	1.8436	2.0731	2.3737
45	0.2439	0.4231	0.4884	0.5351	0.5735	0.6071	0.6375	1.4627	1.5272	1.6046	1.7026	1.8381	2.0656	2.3633
46	0.2445	0.4237	0.4890	0.5357	0.5741	0.6076	0.6380	1.4600	1.5241	1.6011	1.6984	1.8329	2.0585	2.3533
47	0.2450	0.4244	0.4896	0.5363	0.5746	0.6081	0.6385	1.4575	1.5213	1.5977	1.6944	1.8279	2.0517	2.3439
48	0.2455	0.4250	0.4902	0.5369	0.5752	0.6086	0.6389	1.4551	1.5185	1.5945	1.6905	1.8231	2.0452	2.3348
49	0.2459	0.4255	0.4908	0.5374	0.5757	0.6091	0.6394	1.4528	1.5158	1.5914	1.6868	1.8185	2.0389	2.3261
50	0.2464	0.4261	0.4913	0.5379	0.5762	0.6096	0.6398	1.4506	1.5133	1.5884	1.6833	1.8141	2.0330	2.3178

Tabla de Cuantiles de la Distribución F

Grados de libertad del numerador en extremo superior izquierdo, grados de libertad del denominador en margen izquierdo de cada fila. En el margen superior se lee P(F ≤ x) para los valores de x en el cuerpo de la tabla.

19	0.001	0.025	0.050	0.075	0.100	0.125	0.150	0.850	0.875	0.900	0.925	0.950	0.975	0.990
1	0.0663	0.1689	0.2283	0.2818	0.3345	0.3882	0.4443	27.2075	39.3343	61.6578	109.887	247.688	991.800	6200.75
2	0.0985	0.2219	0.2839	0.3358	0.3838	0.4302	0.4762	6.1006	7.4364	9.4387	12.7743	19.4432	39.4457	99.4478
3	0.1208	0.2562	0.3198	0.3711	0.4172	0.4608	0.5031	3.7786	4.3658	5.1870	6.4408	8.6670	14.1808	26.7191
4	0.1376	0.2810	0.3454	0.3963	0.4412	0.4831	0.5233	2.9725	3.3454	3.8485	4.5831	5.8114	8.5753	14.0481
5	0.1510	0.3001	0.3650	0.4154	0.4596	0.5002	0.5389	2.5690	2.8461	3.2117	3.7310	4.5678	6.3444	9.5797
6	0.1619	0.3153	0.3805	0.4306	0.4741	0.5138	0.5514	2.3273	2.5511	2.8419	3.2474	3.8844	5.1844	7.4219
7	0.1711	0.3278	0.3932	0.4430	0.4859	0.5249	0.5616	2.1661	2.3563	2.6008	2.9368	3.4551	4.4829	6.1808
8	0.1789	0.3383	0.4038	0.4533	0.4958	0.5342	0.5702	2.0508	2.2180	2.4310	2.7207	3.1612	4.0158	5.3841
9	0.1856	0.3472	0.4128	0.4621	0.5041	0.5421	0.5774	1.9641	2.1146	2.3050	2.5617	2.9477	3.6833	4.8327
10	0.1915	0.3550	0.4205	0.4696	0.5113	0.5488	0.5837	1.8966	2.0343	2.2077	2.4397	2.7854	3.4351	4.4299
11	0.1967	0.3617	0.4273	0.4762	0.5176	0.5547	0.5891	1.8423	1.9701	2.1302	2.3432	2.6581	3.2428	4.1234
12	0.2013	0.3677	0.4333	0.4820	0.5231	0.5599	0.5939	1.7978	1.9176	2.0670	2.2648	2.5554	3.0896	3.8827
13	0.2055	0.3730	0.4386	0.4871	0.5280	0.5645	0.5982	1.7606	1.8738	2.0145	2.2000	2.4709	2.9646	3.6889
14	0.2092	0.3778	0.4433	0.4917	0.5323	0.5686	0.6020	1.7290	1.8367	1.9701	2.1454	2.4000	2.8607	3.5294
15	0.2126	0.3821	0.4476	0.4958	0.5363	0.5723	0.6054	1.7018	1.8048	1.9321	2.0987	2.3398	2.7730	3.3961
16	0.2157	0.3860	0.4515	0.4996	0.5398	0.5756	0.6085	1.6782	1.7772	1.8992	2.0585	2.2880	2.6980	3.2829
17	0.2185	0.3896	0.4550	0.5030	0.5431	0.5786	0.6113	1.6575	1.7530	1.8704	2.0233	2.2429	2.6331	3.1857
18	0.2211	0.3928	0.4582	0.5061	0.5460	0.5814	0.6139	1.6391	1.7316	1.8450	1.9923	2.2033	2.5764	3.1013
19	0.2235	0.3958	0.4612	0.5089	0.5487	0.5839	0.6162	1.6228	1.7125	1.8224	1.9649	2.1682	2.5264	3.0274
20	0.2258	0.3986	0.4639	0.5116	0.5512	0.5863	0.6184	1.6081	1.6954	1.8022	1.9403	2.1370	2.4821	2.9620
21	0.2278	0.4011	0.4665	0.5140	0.5535	0.5884	0.6204	1.5949	1.6800	1.7840	1.9182	2.1090	2.4424	2.9038
22	0.2297	0.4035	0.4688	0.5163	0.5557	0.5905	0.6223	1.5829	1.6661	1.7675	1.8983	2.0837	2.4067	2.8518
23	0.2315	0.4057	0.4710	0.5183	0.5576	0.5923	0.6240	1.5719	1.6533	1.7525	1.8802	2.0608	2.3745	2.8049
24	0.2332	0.4078	0.4730	0.5203	0.5595	0.5941	0.6256	1.5619	1.6417	1.7388	1.8636	2.0399	2.3452	2.7624
25	0.2348	0.4097	0.4749	0.5221	0.5612	0.5957	0.6271	1.5527	1.6311	1.7263	1.8485	2.0207	2.3184	2.7238
26	0.2363	0.4115	0.4767	0.5238	0.5629	0.5972	0.6285	1.5443	1.6212	1.7147	1.8345	2.0032	2.2939	2.6885
27	0.2376	0.4132	0.4784	0.5254	0.5644	0.5986	0.6298	1.5364	1.6121	1.7040	1.8217	1.9870	2.2713	2.6561
28	0.2390	0.4148	0.4799	0.5269	0.5658	0.6000	0.6311	1.5291	1.6037	1.6941	1.8097	1.9720	2.2505	2.6263
29	0.2402	0.4163	0.4814	0.5284	0.5672	0.6012	0.6323	1.5224	1.5959	1.6849	1.7987	1.9581	2.2313	2.5987
30	0.2414	0.4178	0.4828	0.5297	0.5684	0.6024	0.6334	1.5161	1.5886	1.6763	1.7884	1.9452	2.2134	2.5732
31	0.2425	0.4191	0.4841	0.5310	0.5696	0.6036	0.6344	1.5102	1.5818	1.6683	1.7788	1.9332	2.1967	2.5496
32	0.2435	0.4204	0.4854	0.5322	0.5708	0.6046	0.6354	1.5046	1.5754	1.6608	1.7698	1.9219	2.1812	2.5275
33	0.2446	0.4216	0.4866	0.5333	0.5719	0.6056	0.6363	1.4994	1.5694	1.6538	1.7614	1.9114	2.1667	2.5069
34	0.2455	0.4228	0.4877	0.5344	0.5729	0.6066	0.6372	1.4945	1.5637	1.6472	1.7534	1.9015	2.1531	2.4876
35	0.2464	0.4239	0.4888	0.5354	0.5739	0.6075	0.6381	1.4899	1.5584	1.6410	1.7460	1.8922	2.1403	2.4695
36	0.2473	0.4249	0.4898	0.5364	0.5748	0.6084	0.6389	1.4856	1.5534	1.6351	1.7389	1.8834	2.1282	2.4526
37	0.2481	0.4259	0.4908	0.5374	0.5757	0.6092	0.6397	1.4815	1.5486	1.6296	1.7323	1.8752	2.1168	2.4366
38	0.2489	0.4269	0.4917	0.5383	0.5765	0.6100	0.6404	1.4776	1.5442	1.6243	1.7260	1.8673	2.1061	2.4215
39	0.2497	0.4278	0.4926	0.5391	0.5773	0.6108	0.6411	1.4739	1.5399	1.6193	1.7201	1.8599	2.0960	2.4072
40	0.2504	0.4286	0.4935	0.5399	0.5781	0.6115	0.6418	1.4704	1.5358	1.6146	1.7144	1.8529	2.0864	2.3937
41	0.2511	0.4295	0.4943	0.5407	0.5788	0.6122	0.6424	1.4671	1.5320	1.6101	1.7091	1.8462	2.0772	2.3810
42	0.2518	0.4303	0.4951	0.5415	0.5796	0.6129	0.6430	1.4639	1.5283	1.6058	1.7039	1.8399	2.0686	2.3688
43	0.2524	0.4311	0.4958	0.5422	0.5802	0.6135	0.6436	1.4608	1.5248	1.6017	1.6991	1.8338	2.0603	2.3573
44	0.2530	0.4318	0.4966	0.5429	0.5809	0.6141	0.6442	1.4580	1.5215	1.5979	1.6944	1.8281	2.0525	2.3464
45	0.2536	0.4325	0.4973	0.5435	0.5815	0.6147	0.6448	1.4552	1.5183	1.5941	1.6900	1.8226	2.0450	2.3359
46	0.2542	0.4332	0.4979	0.5442	0.5821	0.6153	0.6453	1.4526	1.5153	1.5906	1.6858	1.8173	2.0379	2.3259
47	0.2547	0.4338	0.4986	0.5448	0.5827	0.6158	0.6458	1.4500	1.5124	1.5872	1.6817	1.8123	2.0310	2.3164
48	0.2553	0.4345	0.4992	0.5454	0.5833	0.6163	0.6463	1.4476	1.5096	1.5839	1.6779	1.8075	2.0245	2.3073
49	0.2558	0.4351	0.4998	0.5460	0.5838	0.6168	0.6467	1.4453	1.5069	1.5808	1.6741	1.8029	2.0182	2.2986
50	0.2563	0.4357	0.5004	0.5465	0.5843	0.6173	0.6472	1.4430	1.5043	1.5778	1.6706	1.7985	2.0122	2.2903

Tabla de Cuantiles de la Distribución F

Grados de libertad del numerador en extremo superior izquierdo, grados de libertad del denominador en margen izquierdo de cada fila. En el margen superior se lee P(F ≤ x) para los valores de x en el cuerpo de la tabla.

20	0.001	0.025	0.050	0.075	0.100	0.125	0.150	0.850	0.875	0.900	0.925	0.950	0.975	0.990
1	0.0675	0.1703	0.2298	0.2834	0.3362	0.3900	0.4462	27.2448	39.3875	61.7401	110.032	248.016	993.081	6208.66
2	0.1005	0.2242	0.2863	0.3382	0.3862	0.4326	0.4787	6.1033	7.4390	9.4413	12.7769	19.4457	39.4475	99.4478
3	0.1235	0.2592	0.3227	0.3740	0.4202	0.4637	0.5060	3.7778	4.3643	5.1845	6.4367	8.6602	14.1674	26.6900
4	0.1409	0.2845	0.3489	0.3997	0.4447	0.4865	0.5266	2.9704	3.3424	3.8443	4.5772	5.8025	8.5599	14.0194
5	0.1548	0.3040	0.3689	0.4193	0.4633	0.5039	0.5425	2.5662	2.8423	3.2067	3.7242	4.5581	6.3285	9.5527
6	0.1662	0.3197	0.3848	0.4348	0.4782	0.5178	0.5553	2.3239	2.5467	2.8363	3.2400	3.8742	5.1684	7.3958
7	0.1757	0.3325	0.3978	0.4475	0.4903	0.5292	0.5657	2.1623	2.3516	2.5947	2.9290	3.4445	4.4668	6.1555
8	0.1838	0.3433	0.4087	0.4581	0.5004	0.5387	0.5745	2.0468	2.2129	2.4246	2.7125	3.1503	3.9994	5.3591
9	0.1909	0.3525	0.4179	0.4671	0.5089	0.5467	0.5819	1.9598	2.1093	2.2983	2.5532	2.9365	3.6669	4.8080
10	0.1970	0.3605	0.4259	0.4748	0.5163	0.5536	0.5883	1.8920	2.0288	2.2007	2.4310	2.7740	3.4185	4.4054
11	0.2025	0.3675	0.4329	0.4816	0.5228	0.5597	0.5939	1.8376	1.9644	2.1230	2.3343	2.6464	3.2261	4.0990
12	0.2073	0.3737	0.4391	0.4875	0.5284	0.5650	0.5988	1.7929	1.9116	2.0597	2.2557	2.5436	3.0728	3.8584
13	0.2117	0.3792	0.4445	0.4928	0.5335	0.5697	0.6032	1.7556	1.8677	2.0070	2.1907	2.4589	2.9477	3.6646
14	0.2156	0.3842	0.4494	0.4976	0.5380	0.5740	0.6071	1.7238	1.8304	1.9625	2.1359	2.3879	2.8437	3.5052
15	0.2192	0.3886	0.4539	0.5018	0.5420	0.5778	0.6107	1.6965	1.7984	1.9243	2.0891	2.3275	2.7559	3.3719
16	0.2225	0.3927	0.4579	0.5057	0.5457	0.5812	0.6139	1.6728	1.7707	1.8913	2.0487	2.2756	2.6808	3.2587
17	0.2255	0.3964	0.4615	0.5092	0.5490	0.5843	0.6168	1.6520	1.7463	1.8624	2.0134	2.2304	2.6158	3.1615
18	0.2282	0.3998	0.4649	0.5124	0.5521	0.5872	0.6194	1.6335	1.7248	1.8368	1.9824	2.1906	2.5590	3.0771
19	0.2308	0.4029	0.4679	0.5154	0.5549	0.5898	0.6218	1.6171	1.7057	1.8142	1.9548	2.1555	2.5089	3.0031
20	0.2331	0.4058	0.4708	0.5181	0.5575	0.5922	0.6241	1.6023	1.6885	1.7938	1.9301	2.1242	2.4645	2.9377
21	0.2353	0.4084	0.4734	0.5206	0.5598	0.5945	0.6262	1.5890	1.6730	1.7756	1.9080	2.0960	2.4247	2.8795
22	0.2373	0.4109	0.4758	0.5230	0.5621	0.5966	0.6281	1.5769	1.6590	1.7590	1.8879	2.0707	2.3890	2.8274
23	0.2392	0.4132	0.4781	0.5251	0.5641	0.5985	0.6299	1.5659	1.6462	1.7439	1.8697	2.0476	2.3566	2.7805
24	0.2410	0.4154	0.4802	0.5272	0.5660	0.6003	0.6315	1.5559	1.6345	1.7302	1.8531	2.0267	2.3273	2.7380
25	0.2427	0.4174	0.4822	0.5290	0.5678	0.6020	0.6331	1.5466	1.6238	1.7175	1.8379	2.0075	2.3005	2.6993
26	0.2442	0.4193	0.4840	0.5308	0.5695	0.6035	0.6346	1.5381	1.6139	1.7059	1.8238	1.9898	2.2759	2.6640
27	0.2457	0.4210	0.4858	0.5325	0.5711	0.6050	0.6359	1.5302	1.6047	1.6951	1.8109	1.9736	2.2533	2.6316
28	0.2471	0.4227	0.4874	0.5341	0.5726	0.6064	0.6372	1.5228	1.5962	1.6852	1.7989	1.9586	2.2324	2.6018
29	0.2484	0.4243	0.4890	0.5355	0.5740	0.6077	0.6384	1.5160	1.5884	1.6759	1.7878	1.9446	2.2131	2.5742
30	0.2497	0.4258	0.4904	0.5369	0.5753	0.6090	0.6396	1.5097	1.5810	1.6673	1.7775	1.9317	2.1952	2.5487
31	0.2509	0.4272	0.4918	0.5383	0.5766	0.6101	0.6407	1.5037	1.5741	1.6593	1.7678	1.9196	2.1785	2.5249
32	0.2520	0.4285	0.4931	0.5395	0.5777	0.6112	0.6417	1.4982	1.5677	1.6517	1.7588	1.9083	2.1629	2.5028
33	0.2531	0.4298	0.4943	0.5407	0.5789	0.6123	0.6427	1.4929	1.5617	1.6446	1.7503	1.8977	2.1483	2.4822
34	0.2541	0.4310	0.4955	0.5418	0.5799	0.6133	0.6436	1.4880	1.5560	1.6380	1.7423	1.8877	2.1346	2.4629
35	0.2551	0.4322	0.4967	0.5429	0.5809	0.6143	0.6445	1.4834	1.5506	1.6317	1.7348	1.8784	2.1218	2.4448
36	0.2560	0.4333	0.4977	0.5439	0.5819	0.6152	0.6453	1.4790	1.5456	1.6258	1.7278	1.8696	2.1097	2.4278
37	0.2569	0.4343	0.4987	0.5449	0.5828	0.6160	0.6461	1.4748	1.5408	1.6202	1.7211	1.8612	2.0983	2.4118
38	0.2577	0.4353	0.4997	0.5458	0.5837	0.6168	0.6469	1.4709	1.5363	1.6149	1.7147	1.8534	2.0875	2.3967
39	0.2585	0.4363	0.5007	0.5467	0.5846	0.6176	0.6476	1.4672	1.5320	1.6099	1.7088	1.8459	2.0774	2.3824
40	0.2593	0.4372	0.5015	0.5476	0.5854	0.6184	0.6483	1.4637	1.5279	1.6052	1.7031	1.8389	2.0677	2.3689
41	0.2601	0.4381	0.5024	0.5484	0.5861	0.6191	0.6490	1.4603	1.5240	1.6006	1.6977	1.8321	2.0586	2.3561
42	0.2608	0.4389	0.5032	0.5492	0.5869	0.6198	0.6496	1.4571	1.5203	1.5963	1.6925	1.8258	2.0499	2.3439
43	0.2615	0.4397	0.5040	0.5499	0.5876	0.6205	0.6503	1.4540	1.5168	1.5922	1.6876	1.8197	2.0416	2.3324
44	0.2621	0.4405	0.5048	0.5507	0.5883	0.6211	0.6509	1.4511	1.5134	1.5883	1.6830	1.8139	2.0337	2.3214
45	0.2628	0.4412	0.5055	0.5513	0.5889	0.6217	0.6514	1.4483	1.5102	1.5846	1.6785	1.8084	2.0262	2.3109
46	0.2634	0.4420	0.5062	0.5520	0.5896	0.6223	0.6520	1.4457	1.5072	1.5810	1.6742	1.8031	2.0190	2.3009
47	0.2640	0.4427	0.5069	0.5527	0.5902	0.6229	0.6525	1.4431	1.5042	1.5776	1.6702	1.7980	2.0122	2.2914
48	0.2645	0.4433	0.5075	0.5533	0.5907	0.6234	0.6530	1.4407	1.5014	1.5743	1.6663	1.7932	2.0056	2.2823
49	0.2651	0.4440	0.5081	0.5539	0.5913	0.6240	0.6535	1.4383	1.4987	1.5711	1.6625	1.7886	1.9993	2.2736
50	0.2656	0.4446	0.5087	0.5544	0.5919	0.6245	0.6540	1.4361	1.4961	1.5681	1.6589	1.7841	1.9933	2.2652

Tabla de Cuantiles de la Distribución F

Grados de libertad del numerador en extremo superior izquierdo, grados de libertad del denominador en margen izquierdo de cada fila. En el margen superior se lee P(F ≤ x) para los valores de x en el cuerpo de la tabla.

25	0.001	0.025	0.050	0.075	0.100	0.125	0.150	0.850	0.875	0.900	0.925	0.950	0.975	0.990
1	0.0721	0.1759	0.2358	0.2897	0.3427	0.3969	0.4534	27.3862	39.5896	62.0548	110.589	249.260	998.087	6239.86
2	0.1084	0.2330	0.2954	0.3474	0.3955	0.4420	0.4881	6.1132	7.4489	9.4513	12.7869	19.4557	39.4575	99.4587
3	0.1342	0.2707	0.3343	0.3855	0.4316	0.4750	0.5172	3.7746	4.3583	5.1747	6.4211	8.6341	14.1154	26.5791
4	0.1540	0.2982	0.3625	0.4131	0.4578	0.4994	0.5393	2.9623	3.3309	3.8283	4.5545	5.7687	8.5010	13.9107
5	0.1699	0.3196	0.3842	0.4343	0.4780	0.5182	0.5564	2.5552	2.8277	3.1873	3.6980	4.5209	6.2678	9.4492
6	0.1831	0.3369	0.4015	0.4511	0.4941	0.5333	0.5703	2.3110	2.5300	2.8147	3.2115	3.8348	5.1069	7.2960
7	0.1942	0.3511	0.4158	0.4650	0.5073	0.5456	0.5817	2.1479	2.3332	2.5714	2.8987	3.4036	4.4045	6.0579
8	0.2038	0.3632	0.4279	0.4767	0.5183	0.5560	0.5912	2.0310	2.1933	2.3999	2.6809	3.1081	3.9367	5.2631
9	0.2122	0.3736	0.4382	0.4866	0.5278	0.5649	0.5994	1.9431	2.0885	2.2725	2.5204	2.8932	3.6035	4.7130
10	0.2195	0.3826	0.4471	0.4952	0.5360	0.5725	0.6065	1.8744	2.0071	2.1739	2.3972	2.7298	3.3546	4.3111
11	0.2261	0.3906	0.4550	0.5028	0.5431	0.5792	0.6127	1.8192	1.9418	2.0953	2.2995	2.6014	3.1616	4.0051
12	0.2319	0.3976	0.4619	0.5094	0.5494	0.5852	0.6182	1.7738	1.8884	2.0312	2.2202	2.4977	3.0077	3.7647
13	0.2372	0.4039	0.4681	0.5154	0.5551	0.5904	0.6231	1.7358	1.8437	1.9778	2.1544	2.4123	2.8821	3.5710
14	0.2420	0.4096	0.4737	0.5207	0.5601	0.5952	0.6274	1.7036	1.8059	1.9326	2.0990	2.3407	2.7777	3.4116
15	0.2464	0.4148	0.4787	0.5256	0.5647	0.5995	0.6314	1.6757	1.7733	1.8939	2.0516	2.2797	2.6894	3.2782
16	0.2504	0.4195	0.4833	0.5299	0.5689	0.6033	0.6350	1.6515	1.7451	1.8603	2.0106	2.2272	2.6138	3.1650
17	0.2541	0.4238	0.4875	0.5339	0.5726	0.6069	0.6383	1.6303	1.7203	1.8309	1.9748	2.1815	2.5484	3.0676
18	0.2575	0.4277	0.4913	0.5376	0.5761	0.6101	0.6413	1.6114	1.6983	1.8049	1.9433	2.1413	2.4912	2.9831
19	0.2606	0.4313	0.4949	0.5410	0.5793	0.6131	0.6440	1.5946	1.6788	1.7818	1.9153	2.1057	2.4408	2.9089
20	0.2636	0.4347	0.4981	0.5441	0.5822	0.6159	0.6466	1.5795	1.6612	1.7611	1.8902	2.0739	2.3959	2.8434
21	0.2663	0.4378	0.5012	0.5470	0.5850	0.6184	0.6489	1.5659	1.6454	1.7424	1.8676	2.0454	2.3558	2.7850
22	0.2688	0.4407	0.5040	0.5497	0.5875	0.6208	0.6511	1.5535	1.6310	1.7255	1.8472	2.0196	2.3198	2.7328
23	0.2712	0.4434	0.5066	0.5522	0.5899	0.6230	0.6532	1.5422	1.6179	1.7101	1.8287	1.9963	2.2871	2.6857
24	0.2734	0.4460	0.5091	0.5545	0.5921	0.6251	0.6551	1.5319	1.6059	1.6960	1.8117	1.9750	2.2574	2.6430
25	0.2755	0.4484	0.5114	0.5567	0.5941	0.6270	0.6569	1.5223	1.5949	1.6831	1.7962	1.9554	2.2303	2.6041
26	0.2775	0.4506	0.5136	0.5588	0.5961	0.6288	0.6586	1.5136	1.5848	1.6712	1.7819	1.9375	2.2054	2.5686
27	0.2794	0.4527	0.5156	0.5607	0.5979	0.6305	0.6601	1.5054	1.5754	1.6602	1.7687	1.9210	2.1826	2.5360
28	0.2812	0.4547	0.5175	0.5626	0.5996	0.6321	0.6616	1.4979	1.5667	1.6500	1.7564	1.9057	2.1614	2.5060
29	0.2829	0.4566	0.5193	0.5643	0.6013	0.6336	0.6630	1.4909	1.5586	1.6405	1.7451	1.8915	2.1419	2.4783
30	0.2844	0.4584	0.5211	0.5659	0.6028	0.6351	0.6644	1.4843	1.5510	1.6316	1.7345	1.8782	2.1237	2.4526
31	0.2860	0.4601	0.5227	0.5675	0.6042	0.6364	0.6656	1.4782	1.5439	1.6234	1.7246	1.8659	2.1068	2.4287
32	0.2874	0.4617	0.5242	0.5689	0.6056	0.6377	0.6668	1.4724	1.5373	1.6156	1.7153	1.8544	2.0910	2.4065
33	0.2888	0.4632	0.5257	0.5703	0.6069	0.6389	0.6679	1.4670	1.5311	1.6083	1.7066	1.8436	2.0762	2.3857
34	0.2901	0.4646	0.5271	0.5717	0.6082	0.6401	0.6690	1.4620	1.5252	1.6015	1.6984	1.8334	2.0623	2.3662
35	0.2913	0.4660	0.5284	0.5729	0.6094	0.6412	0.6700	1.4572	1.5197	1.5950	1.6908	1.8239	2.0493	2.3480
36	0.2925	0.4674	0.5297	0.5741	0.6105	0.6423	0.6710	1.4526	1.5145	1.5890	1.6835	1.8149	2.0370	2.3308
37	0.2937	0.4686	0.5309	0.5753	0.6116	0.6433	0.6720	1.4483	1.5095	1.5832	1.6766	1.8064	2.0254	2.3147
38	0.2948	0.4698	0.5321	0.5764	0.6126	0.6442	0.6729	1.4443	1.5049	1.5778	1.6701	1.7983	2.0145	2.2994
39	0.2958	0.4710	0.5332	0.5774	0.6136	0.6452	0.6737	1.4404	1.5004	1.5726	1.6640	1.7907	2.0042	2.2850
40	0.2968	0.4721	0.5342	0.5784	0.6146	0.6460	0.6745	1.4368	1.4962	1.5677	1.6581	1.7835	1.9943	2.2714
41	0.2978	0.4731	0.5353	0.5794	0.6155	0.6469	0.6753	1.4333	1.4922	1.5630	1.6526	1.7766	1.9850	2.2585
42	0.2987	0.4742	0.5362	0.5803	0.6163	0.6477	0.6761	1.4300	1.4884	1.5586	1.6473	1.7701	1.9762	2.2462
43	0.2996	0.4751	0.5372	0.5812	0.6172	0.6485	0.6768	1.4268	1.4848	1.5543	1.6423	1.7638	1.9678	2.2345
44	0.3004	0.4761	0.5381	0.5821	0.6180	0.6492	0.6775	1.4238	1.4813	1.5503	1.6375	1.7579	1.9597	2.2234
45	0.3013	0.4770	0.5390	0.5829	0.6188	0.6500	0.6781	1.4209	1.4780	1.5464	1.6329	1.7522	1.9521	2.2129
46	0.3021	0.4779	0.5398	0.5837	0.6195	0.6507	0.6788	1.4181	1.4748	1.5427	1.6285	1.7468	1.9448	2.2028
47	0.3028	0.4787	0.5406	0.5845	0.6202	0.6513	0.6794	1.4154	1.4717	1.5392	1.6243	1.7416	1.9378	2.1931
48	0.3036	0.4795	0.5414	0.5852	0.6209	0.6520	0.6800	1.4129	1.4688	1.5358	1.6203	1.7367	1.9311	2.1839
49	0.3043	0.4803	0.5421	0.5859	0.6216	0.6526	0.6806	1.4105	1.4660	1.5325	1.6164	1.7319	1.9247	2.1751
50	0.3050	0.4811	0.5429	0.5866	0.6222	0.6532	0.6811	1.4081	1.4633	1.5294	1.6127	1.7273	1.9186	2.1667

Tabla de Cuantiles de la Distribución F

Grados de libertad del numerador en extremo superior izquierdo, grados de libertad del denominador en margen izquierdo de cada fila. En el margen superior se lee P(F ≤ x) para los valores de x en el cuerpo de la tabla.

30	0.001	0.025	0.050	0.075	0.100	0.125	0.150	0.850	0.875	0.900	0.925	0.950	0.975	0.990
1	0.0752	0.1796	0.2398	0.2939	0.3471	0.4015	0.4582	27.4811	39.7251	62.2649	110.962	250.096	1001.40	6260.35
2	0.1140	0.2391	0.3016	0.3537	0.4018	0.4483	0.4945	6.1198	7.4556	9.4579	12.7935	19.4625	39.4648	99.4660
3	0.1418	0.2786	0.3422	0.3934	0.4394	0.4827	0.5248	3.7724	4.3543	5.1681	6.4105	8.6166	14.0806	26.5045
4	0.1633	0.3077	0.3718	0.4223	0.4668	0.5082	0.5479	2.9567	3.3231	3.8174	4.5392	5.7459	8.4613	13.8375
5	0.1807	0.3304	0.3947	0.4445	0.4880	0.5280	0.5659	2.5477	2.8177	3.1741	3.6802	4.4957	6.2269	9.3794
6	0.1952	0.3488	0.4131	0.4624	0.5050	0.5439	0.5805	2.3021	2.5186	2.8000	3.1921	3.8082	5.0652	7.2286
7	0.2076	0.3642	0.4284	0.4771	0.5190	0.5570	0.5926	2.1379	2.3207	2.5555	2.8782	3.3758	4.3624	5.9920
8	0.2183	0.3772	0.4413	0.4895	0.5308	0.5680	0.6028	2.0202	2.1798	2.3830	2.6593	3.0794	3.8940	5.1981
9	0.2276	0.3884	0.4523	0.5002	0.5408	0.5774	0.6115	1.9316	2.0742	2.2547	2.4980	2.8637	3.5604	4.6486
10	0.2359	0.3982	0.4620	0.5094	0.5496	0.5856	0.6191	1.8622	1.9921	2.1554	2.3740	2.6996	3.3110	4.2469
11	0.2433	0.4069	0.4705	0.5176	0.5573	0.5928	0.6257	1.8065	1.9263	2.0762	2.2757	2.5705	3.1176	3.9411
12	0.2500	0.4146	0.4780	0.5248	0.5641	0.5992	0.6316	1.7606	1.8723	2.0115	2.1958	2.4663	2.9633	3.7008
13	0.2560	0.4215	0.4847	0.5312	0.5702	0.6049	0.6369	1.7221	1.8272	1.9576	2.1295	2.3803	2.8373	3.5070
14	0.2615	0.4278	0.4908	0.5370	0.5757	0.6100	0.6416	1.6894	1.7889	1.9119	2.0735	2.3082	2.7324	3.3476
15	0.2665	0.4334	0.4963	0.5423	0.5806	0.6146	0.6459	1.6613	1.7559	1.8728	2.0257	2.2468	2.6437	3.2141
16	0.2711	0.4386	0.5013	0.5471	0.5851	0.6188	0.6498	1.6367	1.7273	1.8388	1.9843	2.1938	2.5678	3.1007
17	0.2753	0.4434	0.5059	0.5514	0.5893	0.6227	0.6533	1.6151	1.7021	1.8090	1.9481	2.1477	2.5020	3.0032
18	0.2792	0.4477	0.5102	0.5555	0.5930	0.6262	0.6566	1.5960	1.6798	1.7827	1.9162	2.1071	2.4445	2.9185
19	0.2829	0.4518	0.5141	0.5592	0.5965	0.6295	0.6596	1.5789	1.6600	1.7592	1.8878	2.0712	2.3937	2.8442
20	0.2863	0.4555	0.5177	0.5626	0.5998	0.6325	0.6624	1.5635	1.6421	1.7382	1.8624	2.0391	2.3486	2.7785
21	0.2895	0.4590	0.5210	0.5658	0.6028	0.6353	0.6650	1.5496	1.6260	1.7193	1.8396	2.0102	2.3082	2.7200
22	0.2925	0.4623	0.5242	0.5688	0.6056	0.6379	0.6674	1.5370	1.6114	1.7021	1.8189	1.9842	2.2718	2.6675
23	0.2952	0.4653	0.5271	0.5715	0.6082	0.6403	0.6696	1.5255	1.5981	1.6864	1.8000	1.9605	2.2389	2.6202
24	0.2979	0.4682	0.5298	0.5741	0.6106	0.6426	0.6717	1.5149	1.5859	1.6721	1.7828	1.9390	2.2090	2.5773
25	0.3003	0.4709	0.5324	0.5765	0.6129	0.6447	0.6737	1.5052	1.5746	1.6589	1.7670	1.9192	2.1816	2.5383
26	0.3027	0.4734	0.5348	0.5788	0.6150	0.6467	0.6756	1.4962	1.5643	1.6468	1.7525	1.9010	2.1565	2.5026
27	0.3049	0.4758	0.5371	0.5810	0.6171	0.6486	0.6773	1.4879	1.5547	1.6356	1.7391	1.8842	2.1334	2.4699
28	0.3070	0.4780	0.5393	0.5830	0.6190	0.6504	0.6789	1.4802	1.5458	1.6252	1.7266	1.8687	2.1121	2.4397
29	0.3090	0.4802	0.5413	0.5849	0.6208	0.6521	0.6805	1.4730	1.5375	1.6155	1.7150	1.8543	2.0923	2.4118
30	0.3108	0.4822	0.5432	0.5868	0.6225	0.6537	0.6820	1.4663	1.5298	1.6065	1.7042	1.8409	2.0739	2.3860
31	0.3126	0.4841	0.5451	0.5885	0.6241	0.6552	0.6834	1.4601	1.5226	1.5980	1.6942	1.8283	2.0568	2.3619
32	0.3143	0.4859	0.5468	0.5901	0.6256	0.6566	0.6847	1.4542	1.5158	1.5901	1.6847	1.8166	2.0408	2.3395
33	0.3160	0.4876	0.5484	0.5917	0.6271	0.6580	0.6860	1.4486	1.5094	1.5827	1.6759	1.8056	2.0259	2.3186
34	0.3175	0.4893	0.5500	0.5932	0.6285	0.6593	0.6872	1.4434	1.5034	1.5757	1.6675	1.7953	2.0118	2.2990
35	0.3190	0.4909	0.5515	0.5946	0.6298	0.6605	0.6883	1.4385	1.4977	1.5691	1.6597	1.7856	1.9986	2.2806
36	0.3204	0.4924	0.5530	0.5960	0.6311	0.6617	0.6894	1.4338	1.4924	1.5629	1.6523	1.7764	1.9862	2.2633
37	0.3218	0.4938	0.5543	0.5973	0.6323	0.6628	0.6904	1.4294	1.4873	1.5570	1.6453	1.7678	1.9745	2.2470
38	0.3231	0.4952	0.5556	0.5985	0.6335	0.6639	0.6914	1.4253	1.4825	1.5514	1.6386	1.7596	1.9634	2.2317
39	0.3243	0.4965	0.5569	0.5997	0.6346	0.6649	0.6924	1.4213	1.4780	1.5461	1.6323	1.7518	1.9529	2.2171
40	0.3256	0.4978	0.5581	0.6008	0.6356	0.6659	0.6933	1.4175	1.4737	1.5411	1.6264	1.7444	1.9429	2.2034
41	0.3267	0.4990	0.5593	0.6019	0.6367	0.6669	0.6942	1.4139	1.4695	1.5363	1.6207	1.7374	1.9335	2.1903
42	0.3278	0.5002	0.5604	0.6029	0.6376	0.6678	0.6950	1.4105	1.4656	1.5317	1.6153	1.7308	1.9245	2.1780
43	0.3289	0.5013	0.5614	0.6039	0.6386	0.6687	0.6958	1.4073	1.4619	1.5274	1.6101	1.7244	1.9159	2.1662
44	0.3299	0.5024	0.5625	0.6049	0.6395	0.6695	0.6966	1.4042	1.4583	1.5232	1.6052	1.7184	1.9078	2.1550
45	0.3309	0.5034	0.5635	0.6058	0.6404	0.6703	0.6973	1.4012	1.4549	1.5193	1.6005	1.7126	1.9000	2.1443
46	0.3319	0.5044	0.5644	0.6067	0.6412	0.6711	0.6981	1.3983	1.4516	1.5155	1.5960	1.7070	1.8926	2.1341
47	0.3328	0.5054	0.5653	0.6076	0.6420	0.6719	0.6988	1.3956	1.4485	1.5118	1.5917	1.7017	1.8855	2.1244
48	0.3337	0.5063	0.5662	0.6084	0.6428	0.6726	0.6994	1.3930	1.4455	1.5084	1.5876	1.6967	1.8787	2.1150
49	0.3346	0.5072	0.5671	0.6093	0.6435	0.6733	0.7001	1.3905	1.4426	1.5050	1.5836	1.6918	1.8722	2.1061
50	0.3354	0.5081	0.5679	0.6100	0.6443	0.6740	0.7007	1.3880	1.4398	1.5018	1.5798	1.6872	1.8659	2.0976

Tabla de Cuantiles de la Distribución F

Grados de libertad del numerador en extremo superior izquierdo, grados de libertad del denominador en margen izquierdo de cada fila. En el margen superior se lee P(F ≤ x) para los valores de x en el cuerpo de la tabla.

40	0.001	0.025	0.050	0.075	0.100	0.125	0.150	0.850	0.875	0.900	0.925	0.950	0.975	0.990
1	0.0793	0.1844	0.2448	0.2992	0.3527	0.4073	0.4642	27.6000	39.8950	62.5291	111.429	251.144	1005.59	6286.43
2	0.1212	0.2469	0.3094	0.3616	0.4098	0.4563	0.5025	6.1282	7.4639	9.4662	12.8018	19.4707	39.4730	99.4769
3	0.1516	0.2887	0.3523	0.4034	0.4492	0.4925	0.5344	3.7695	4.3492	5.1597	6.3972	8.5944	14.0365	26.4108
4	0.1755	0.3199	0.3837	0.4339	0.4783	0.5194	0.5589	2.9496	3.3131	3.8036	4.5197	5.7170	8.4111	13.7452
5	0.1950	0.3444	0.4083	0.4577	0.5008	0.5404	0.5781	2.5381	2.8050	3.1573	3.6577	4.4638	6.1751	9.2912
6	0.2114	0.3644	0.4281	0.4769	0.5190	0.5574	0.5937	2.2907	2.5040	2.7812	3.1675	3.7743	5.0125	7.1432
7	0.2255	0.3811	0.4446	0.4928	0.5340	0.5715	0.6066	2.1252	2.3046	2.5351	2.8519	3.3404	4.3089	5.9084
8	0.2377	0.3954	0.4587	0.5062	0.5468	0.5834	0.6176	2.0063	2.1625	2.3614	2.6317	3.0428	3.8398	5.1156
9	0.2485	0.4078	0.4708	0.5178	0.5578	0.5937	0.6270	1.9167	2.0559	2.2320	2.4693	2.8259	3.5055	4.5667
10	0.2581	0.4187	0.4814	0.5280	0.5673	0.6026	0.6353	1.8465	1.9728	2.1317	2.3443	2.6609	3.2554	4.1653
11	0.2667	0.4284	0.4908	0.5369	0.5757	0.6105	0.6425	1.7900	1.9062	2.0516	2.2451	2.5309	3.0613	3.8596
12	0.2745	0.4370	0.4991	0.5449	0.5832	0.6174	0.6490	1.7434	1.8515	1.9861	2.1643	2.4259	2.9063	3.6192
13	0.2816	0.4448	0.5066	0.5520	0.5900	0.6237	0.6548	1.7043	1.8057	1.9315	2.0973	2.3392	2.7797	3.4253
14	0.2881	0.4519	0.5134	0.5584	0.5960	0.6294	0.6600	1.6710	1.7667	1.8852	2.0406	2.2663	2.6742	3.2657
15	0.2941	0.4583	0.5196	0.5643	0.6015	0.6345	0.6648	1.6423	1.7332	1.8454	1.9922	2.2043	2.5850	3.1319
16	0.2996	0.4642	0.5253	0.5697	0.6066	0.6392	0.6691	1.6173	1.7040	1.8108	1.9502	2.1507	2.5085	3.0182
17	0.3047	0.4696	0.5305	0.5746	0.6112	0.6435	0.6731	1.5952	1.6784	1.7805	1.9134	2.1040	2.4422	2.9204
18	0.3094	0.4747	0.5353	0.5791	0.6154	0.6475	0.6767	1.5757	1.6556	1.7537	1.8809	2.0629	2.3842	2.8354
19	0.3138	0.4793	0.5397	0.5833	0.6194	0.6511	0.6801	1.5582	1.6353	1.7298	1.8521	2.0264	2.3329	2.7608
20	0.3180	0.4836	0.5438	0.5872	0.6230	0.6545	0.6832	1.5424	1.6171	1.7083	1.8262	1.9938	2.2873	2.6947
21	0.3218	0.4877	0.5477	0.5908	0.6264	0.6577	0.6861	1.5282	1.6006	1.6890	1.8029	1.9645	2.2465	2.6359
22	0.3255	0.4914	0.5512	0.5942	0.6295	0.6606	0.6888	1.5152	1.5856	1.6714	1.7818	1.9380	2.2097	2.5831
23	0.3289	0.4950	0.5546	0.5973	0.6325	0.6633	0.6914	1.5034	1.5719	1.6554	1.7625	1.9139	2.1763	2.5355
24	0.3321	0.4983	0.5577	0.6003	0.6353	0.6659	0.6938	1.4925	1.5594	1.6407	1.7450	1.8920	2.1460	2.4923
25	0.3351	0.5014	0.5607	0.6031	0.6379	0.6683	0.6960	1.4825	1.5479	1.6272	1.7288	1.8718	2.1183	2.4530
26	0.3380	0.5044	0.5635	0.6057	0.6403	0.6706	0.6981	1.4733	1.5372	1.6147	1.7139	1.8533	2.0928	2.4170
27	0.3408	0.5072	0.5661	0.6082	0.6427	0.6728	0.7001	1.4647	1.5274	1.6032	1.7002	1.8361	2.0693	2.3840
28	0.3434	0.5098	0.5686	0.6105	0.6449	0.6748	0.7020	1.4568	1.5182	1.5925	1.6874	1.8203	2.0477	2.3535
29	0.3458	0.5123	0.5710	0.6128	0.6469	0.6768	0.7038	1.4493	1.5096	1.5825	1.6755	1.8055	2.0276	2.3253
30	0.3482	0.5147	0.5733	0.6149	0.6489	0.6786	0.7055	1.4424	1.5017	1.5732	1.6644	1.7918	2.0089	2.2992
31	0.3504	0.5170	0.5754	0.6169	0.6508	0.6803	0.7071	1.4359	1.4942	1.5645	1.6541	1.7790	1.9914	2.2749
32	0.3526	0.5191	0.5774	0.6188	0.6525	0.6820	0.7086	1.4298	1.4872	1.5564	1.6444	1.7670	1.9752	2.2522
33	0.3546	0.5212	0.5794	0.6206	0.6542	0.6835	0.7100	1.4240	1.4806	1.5487	1.6352	1.7557	1.9599	2.2311
34	0.3566	0.5231	0.5812	0.6223	0.6559	0.6850	0.7114	1.4186	1.4744	1.5415	1.6267	1.7451	1.9456	2.2112
35	0.3584	0.5250	0.5830	0.6240	0.6574	0.6865	0.7127	1.4135	1.4685	1.5346	1.6186	1.7351	1.9321	2.1926
36	0.3602	0.5268	0.5847	0.6256	0.6589	0.6878	0.7140	1.4087	1.4630	1.5282	1.6109	1.7257	1.9194	2.1751
37	0.3620	0.5285	0.5863	0.6271	0.6603	0.6892	0.7152	1.4041	1.4577	1.5221	1.6037	1.7168	1.9074	2.1585
38	0.3636	0.5302	0.5878	0.6285	0.6616	0.6904	0.7164	1.3998	1.4527	1.5163	1.5969	1.7084	1.8961	2.1430
39	0.3652	0.5318	0.5893	0.6299	0.6629	0.6916	0.7175	1.3957	1.4480	1.5108	1.5904	1.7004	1.8854	2.1282
40	0.3667	0.5333	0.5907	0.6312	0.6642	0.6928	0.7185	1.3917	1.4435	1.5056	1.5842	1.6928	1.8752	2.1142
41	0.3682	0.5347	0.5921	0.6325	0.6654	0.6939	0.7196	1.3880	1.4392	1.5007	1.5783	1.6856	1.8655	2.1010
42	0.3696	0.5361	0.5934	0.6338	0.6665	0.6949	0.7205	1.3844	1.4351	1.4959	1.5727	1.6787	1.8563	2.0884
43	0.3710	0.5375	0.5947	0.6349	0.6676	0.6960	0.7215	1.3810	1.4312	1.4914	1.5674	1.6722	1.8476	2.0764
44	0.3723	0.5388	0.5959	0.6361	0.6687	0.6969	0.7224	1.3778	1.4275	1.4871	1.5623	1.6659	1.8392	2.0650
45	0.3736	0.5401	0.5971	0.6372	0.6697	0.6979	0.7233	1.3747	1.4239	1.4830	1.5574	1.6599	1.8313	2.0542
46	0.3748	0.5413	0.5982	0.6382	0.6707	0.6988	0.7241	1.3717	1.4205	1.4790	1.5527	1.6542	1.8236	2.0438
47	0.3760	0.5424	0.5993	0.6393	0.6717	0.6997	0.7249	1.3688	1.4173	1.4752	1.5483	1.6488	1.8164	2.0339
48	0.3772	0.5436	0.6004	0.6403	0.6726	0.7006	0.7257	1.3661	1.4141	1.4716	1.5440	1.6435	1.8094	2.0244
49	0.3783	0.5447	0.6014	0.6412	0.6735	0.7014	0.7265	1.3634	1.4111	1.4681	1.5399	1.6385	1.8027	2.0153
50	0.3794	0.5457	0.6024	0.6421	0.6743	0.7022	0.7272	1.3609	1.4082	1.4648	1.5359	1.6337	1.7963	2.0066

Tabla de Cuantiles de la Distribución F

Grados de libertad del numerador en extremo superior izquierdo, grados de libertad del denominador en margen izquierdo de cada fila. En el margen superior se lee P(F ≤ x) para los valores de x en el cuerpo de la tabla.

50	0.001	0.025	0.050	0.075	0.100	0.125	0.150	0.850	0.875	0.900	0.925	0.950	0.975	0.990
1	0.0818	0.1873	0.2479	0.3024	0.3560	0.4108	0.4679	27.6716	39.9973	62.6878	111.710	251.774	1008.09	6302.26
2	0.1257	0.2516	0.3142	0.3664	0.4146	0.4612	0.5074	6.1331	7.4689	9.4713	12.8068	19.4757	39.4775	99.4769
3	0.1578	0.2950	0.3584	0.4094	0.4552	0.4984	0.5403	3.7678	4.3460	5.1546	6.3891	8.5810	14.0099	26.3544
4	0.1832	0.3274	0.3911	0.4411	0.4852	0.5263	0.5655	2.9453	3.3070	3.7952	4.5079	5.6995	8.3808	13.6897
5	0.2040	0.3530	0.4166	0.4658	0.5086	0.5481	0.5855	2.5322	2.7973	3.1471	3.6439	4.4444	6.1436	9.2377
6	0.2216	0.3740	0.4374	0.4858	0.5276	0.5657	0.6017	2.2837	2.4950	2.7697	3.1524	3.7537	4.9804	7.0914
7	0.2368	0.3917	0.4547	0.5024	0.5433	0.5804	0.6152	2.1173	2.2947	2.5226	2.8359	3.3189	4.2763	5.8577
8	0.2501	0.4068	0.4695	0.5166	0.5567	0.5929	0.6267	1.9978	2.1518	2.3481	2.6149	3.0204	3.8067	5.0654
9	0.2619	0.4200	0.4823	0.5288	0.5682	0.6037	0.6366	1.9075	2.0446	2.2180	2.4516	2.8028	3.4719	4.5167
10	0.2724	0.4316	0.4935	0.5395	0.5783	0.6131	0.6453	1.8368	1.9609	2.1171	2.3260	2.6371	3.2214	4.1155
11	0.2819	0.4420	0.5035	0.5490	0.5872	0.6214	0.6530	1.7797	1.8938	2.0364	2.2262	2.5066	3.0268	3.8097
12	0.2905	0.4512	0.5124	0.5574	0.5952	0.6288	0.6598	1.7327	1.8385	1.9704	2.1449	2.4010	2.8714	3.5692
13	0.2983	0.4596	0.5204	0.5650	0.6023	0.6355	0.6660	1.6932	1.7923	1.9153	2.0774	2.3138	2.7443	3.3752
14	0.3055	0.4672	0.5277	0.5719	0.6088	0.6415	0.6715	1.6596	1.7530	1.8686	2.0203	2.2405	2.6384	3.2153
15	0.3122	0.4742	0.5344	0.5782	0.6147	0.6470	0.6766	1.6305	1.7191	1.8284	1.9714	2.1780	2.5488	3.0814
16	0.3183	0.4805	0.5404	0.5839	0.6200	0.6520	0.6812	1.6051	1.6895	1.7934	1.9289	2.1240	2.4719	2.9675
17	0.3240	0.4864	0.5461	0.5892	0.6250	0.6566	0.6854	1.5828	1.6636	1.7628	1.8918	2.0769	2.4053	2.8694
18	0.3293	0.4919	0.5512	0.5941	0.6296	0.6608	0.6894	1.5629	1.6405	1.7356	1.8590	2.0354	2.3468	2.7841
19	0.3343	0.4970	0.5560	0.5986	0.6338	0.6647	0.6930	1.5451	1.6199	1.7114	1.8298	1.9986	2.2952	2.7092
20	0.3389	0.5017	0.5605	0.6028	0.6377	0.6684	0.6963	1.5291	1.6014	1.6896	1.8036	1.9656	2.2493	2.6430
21	0.3433	0.5061	0.5647	0.6067	0.6414	0.6718	0.6995	1.5147	1.5846	1.6700	1.7800	1.9360	2.2081	2.5838
22	0.3474	0.5102	0.5686	0.6104	0.6448	0.6750	0.7024	1.5015	1.5694	1.6521	1.7586	1.9092	2.1710	2.5308
23	0.3513	0.5141	0.5722	0.6138	0.6480	0.6779	0.7052	1.4894	1.5555	1.6358	1.7390	1.8848	2.1374	2.4829
24	0.3549	0.5178	0.5757	0.6170	0.6510	0.6807	0.7077	1.4783	1.5427	1.6209	1.7212	1.8625	2.1067	2.4395
25	0.3584	0.5212	0.5789	0.6201	0.6538	0.6834	0.7102	1.4681	1.5309	1.6072	1.7048	1.8421	2.0787	2.3999
26	0.3617	0.5245	0.5820	0.6230	0.6565	0.6859	0.7125	1.4587	1.5201	1.5945	1.6896	1.8233	2.0530	2.3637
27	0.3648	0.5276	0.5849	0.6257	0.6591	0.6882	0.7146	1.4500	1.5100	1.5827	1.6756	1.8059	2.0293	2.3304
28	0.3678	0.5305	0.5876	0.6282	0.6614	0.6904	0.7167	1.4418	1.5007	1.5718	1.6627	1.7898	2.0073	2.2997
29	0.3706	0.5333	0.5902	0.6307	0.6637	0.6925	0.7186	1.4342	1.4919	1.5617	1.6506	1.7748	1.9870	2.2713
30	0.3733	0.5359	0.5927	0.6330	0.6659	0.6945	0.7204	1.4271	1.4838	1.5522	1.6393	1.7609	1.9681	2.2450
31	0.3759	0.5384	0.5951	0.6352	0.6679	0.6964	0.7222	1.4205	1.4761	1.5433	1.6287	1.7478	1.9504	2.2205
32	0.3784	0.5408	0.5973	0.6373	0.6699	0.6982	0.7239	1.4142	1.4689	1.5349	1.6188	1.7356	1.9339	2.1976
33	0.3807	0.5431	0.5995	0.6393	0.6717	0.7000	0.7255	1.4083	1.4622	1.5271	1.6095	1.7241	1.9184	2.1763
34	0.3830	0.5453	0.6015	0.6412	0.6735	0.7016	0.7270	1.4028	1.4558	1.5197	1.6007	1.7134	1.9039	2.1562
35	0.3852	0.5474	0.6035	0.6430	0.6752	0.7032	0.7284	1.3975	1.4498	1.5127	1.5925	1.7032	1.8902	2.1374
36	0.3872	0.5494	0.6053	0.6448	0.6768	0.7047	0.7298	1.3926	1.4441	1.5061	1.5847	1.6936	1.8773	2.1197
37	0.3892	0.5514	0.6071	0.6464	0.6784	0.7061	0.7311	1.3879	1.4387	1.4999	1.5773	1.6845	1.8652	2.1030
38	0.3912	0.5532	0.6089	0.6480	0.6799	0.7075	0.7324	1.3834	1.4336	1.4939	1.5703	1.6759	1.8536	2.0872
39	0.3930	0.5550	0.6105	0.6496	0.6813	0.7088	0.7336	1.3792	1.4288	1.4883	1.5636	1.6678	1.8427	2.0723
40	0.3948	0.5567	0.6121	0.6511	0.6827	0.7101	0.7348	1.3751	1.4241	1.4830	1.5573	1.6600	1.8324	2.0581
41	0.3965	0.5584	0.6136	0.6525	0.6840	0.7113	0.7359	1.3713	1.4197	1.4779	1.5513	1.6526	1.8225	2.0447
42	0.3982	0.5599	0.6151	0.6539	0.6853	0.7125	0.7370	1.3676	1.4155	1.4730	1.5455	1.6456	1.8132	2.0319
43	0.3998	0.5615	0.6165	0.6552	0.6865	0.7137	0.7381	1.3641	1.4115	1.4684	1.5400	1.6389	1.8043	2.0198
44	0.4013	0.5629	0.6179	0.6565	0.6877	0.7147	0.7391	1.3607	1.4077	1.4639	1.5348	1.6325	1.7958	2.0083
45	0.4028	0.5644	0.6192	0.6577	0.6888	0.7158	0.7400	1.3575	1.4040	1.4597	1.5298	1.6264	1.7876	1.9972
46	0.4043	0.5657	0.6205	0.6589	0.6899	0.7168	0.7410	1.3544	1.4005	1.4556	1.5250	1.6206	1.7799	1.9867
47	0.4057	0.5671	0.6217	0.6600	0.6910	0.7178	0.7419	1.3515	1.3971	1.4517	1.5204	1.6150	1.7724	1.9766
48	0.4070	0.5683	0.6229	0.6611	0.6920	0.7188	0.7428	1.3487	1.3939	1.4480	1.5160	1.6096	1.7653	1.9670
49	0.4083	0.5696	0.6241	0.6622	0.6930	0.7197	0.7436	1.3459	1.3908	1.4444	1.5118	1.6044	1.7585	1.9578
50	0.4096	0.5708	0.6252	0.6632	0.6940	0.7206	0.7444	1.3433	1.3878	1.4409	1.5078	1.5995	1.7520	1.9490

Tabla de Cuantiles de la Distribución F

Grados de libertad del numerador en extremo superior izquierdo, grados de libertad del denominador en margen izquierdo de cada fila. En el margen superior se lee P(F ≤ x) para los valores de x en el cuerpo de la tabla.

ν	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	17.970	26.980	32.820	37.080	40.410	43.120	45.400	47.360	49.070	50.590	51.960	53.200	54.330	55.360
2	6.095	8.331	9.798	10.880	11.740	12.440	13.030	13.540	13.990	14.390	14.750	15.080	15.380	15.650
3	4.501	5.910	6.825	7.502	3.037	8.478	8.853	9.177	9.462	9.717	9.946	10.150	10.350	10.530
4	3.927	5.040	5.757	6.297	6.707	7.053	7.347	7.602	7.926	8.027	8.208	8.373	8.525	8.664
5	3.635	4.602	5.219	5.673	6.033	6.330	6.582	6.802	6.995	7.168	7.324	7.466	7.596	7.717
6	3.461	4.339	4.896	5.305	5.628	5.895	6.122	6.319	6.493	6.649	6.789	6.917	7.034	7.143
7	3.344	4.165	4.681	5.060	5.359	5.606	5.815	5.998	6.158	6.302	6.431	6.550	6.658	6.759
8	3.261	4.041	4.529	4.886	5.167	5.399	5.597	5.767	5.918	6.054	6.175	6.287	6.389	6.483
9	3.199	3.949	4.415	4.756	5.024	5.244	5.432	5.595	5.739	5.867	5.983	6.089	6.186	6.276
10	3.151	3.877	4.327	4.654	4.912	5.124	5.305	5.461	5.599	5.722	5.833	5.935	6.028	6.114
11	3.113	3.820	4.756	4.574	4.823	5.028	5.202	5.353	5.487	5.605	5.713	5.811	5.901	5.984
12	3.082	3.773	4.199	4.508	4.751	4.950	5.119	5.265	5.395	5.511	5.615	5.710	5.798	5.878
13	3.055	3.735	4.151	4.453	4.690	4.885	5.049	5.192	5.318	5.431	5.533	5.625	5.711	5.789
14	3.033	3.702	4.111	4.407	4.639	4.829	4.990	5.131	5.254	5.364	5.463	5.554	5.537	5.714
15	3.014	3.674	4.076	4.367	4.595	4.782	4.940	5.077	5.198	5.306	5.404	5.493	5.574	5.649
16	2.998	3.649	4.046	4.333	4.557	4.741	4.897	5.031	5.150	5.256	5.352	5.439	5.520	5.593
17	2.984	3.628	4.020	4.303	4.524	4.705	4.858	4.991	5.108	5.212	5.307	5.392	5.471	5.544
18	2.971	3.609	3.997	4.277	4.495	4.673	4.824	4.956	5.071	5.174	5.267	5.352	5.429	5.501
19	2.960	3.593	3.977	4.253	4.469	4.645	4.794	4.924	5.038	5.140	5.231	5.315	5.391	5.462
20	2.950	3.578	3.958	4.232	4.445	4.622	4.768	4.896	5.008	5.108	5.199	5.282	5.357	5.427
24	2.919	3.532	3.901	4.166	4.373	4.541	4.684	4.807	4.915	5.012	5.099	5.179	5.251	5.319
30	2.888	3.486	3.345	4.102	4.302	4.464	4.602	4.720	4.824	4.917	5.001	5.077	5.147	5.211
40	2.858	3.442	3.791	4.039	4.232	4.389	4.521	4.635	4.735	4.824	4.904	4.977	5.044	5.106
60	2.829	3.399	3.737	3.977	4.163	4.314	4.441	4.550	4.645	4.732	4.808	4.878	4.942	5.001
120	2.800	3.356	3.685	3.917	4.096	4.241	4.363	4.468	4.560	4.641	4.714	4.781	4.842	4.898
inf	2.772	3.314	3.633	3.858	4.030	4.170	4.286	4.387	4.474	4.552	4.622	4.685	4.743	4.796

ν	16	17	18	19	20	22	24	26	28	30	32	34	36
1	56.320	57.220	58.040	58.430	59.560	60.910	62.120	63.220	64.230	65.150	66.010	66.810	67.560
2	15.910	16.140	16.370	16.570	16.770	17.130	17.450	17.750	18.020	18.270	18.500	18.720	18.920
3	10.690	10.840	10.980	11.110	11.240	11.470	11.680	11.870	12.050	12.210	12.360	12.500	12.630
4	8.794	8.914	9.028	9.134	9.233	9.418	9.584	9.736	9.875	10.000	10.120	10.230	10.340
5	7.828	7.932	8.030	8.122	8.708	8.368	8.512	8.643	8.764	8.875	8.979	9.075	9.165
6	7.244	7.338	7.426	7.508	7.587	7.730	7.861	7.979	8.088	8.189	8.283	8.370	8.452
7	6.852	6.939	7.020	7.097	7.170	7.303	7.423	7.533	7.634	7.728	7.814	7.895	7.972
8	6.571	6.653	6.729	6.802	6.870	6.995	7.109	7.212	7.307	7.395	7.477	7.554	7.625
9	6.359	6.437	6.510	6.579	6.644	6.763	6.871	6.970	7.061	7.145	7.222	7.295	7.363
10	6.194	6.269	6.339	6.405	6.467	6.582	6.686	6.781	6.868	6.948	7.023	7.093	7.159
11	6.062	6.134	6.202	6.265	6.326	6.436	6.536	6.628	6.712	6.790	6.863	6.930	6.994
12	5.953	6.023	6.089	6.151	6.209	6.317	6.414	6.503	6.585	6.560	6.731	6.796	6.358
13	5.862	5.931	5.995	6.055	6.112	6.217	6.312	6.398	6.478	6.551	6.620	6.684	6.744
14	5.786	5.852	5.915	5.974	6.029	6.132	6.224	6.309	6.387	6.459	6.526	6.588	6.647
15	5.720	5.785	5.846	5.904	5.958	6.059	6.149	6.233	6.309	6.379	6.445	6.506	6.554
16	5.662	5.727	5.786	5.843	5.897	5.995	6.084	6.166	6.241	6.310	6.374	6.434	6.491
17	5.612	5.675	5.734	5.790	5.842	5.940	6.027	6.107	6.181	6.249	6.313	6.372	6.427
18	5.563	5.630	5.688	5.743	5.794	5.890	5.977	6.055	6.128	6.195	6.258	6.316	6.371
19	5.528	5.589	5.647	5.701	5.752	5.846	5.932	6.009	6.081	6.147	6.209	6.267	6.321
20	5.493	5.553	5.610	5.663	5.714	5.807	5.891	5.968	6.039	6.104	6.155	6.222	6.275
24	5.381	5.439	5.494	5.545	5.594	5.683	5.764	5.838	5.906	5.968	6.027	6.081	6.132
30	5.271	5.327	5.379	5.429	5.475	5.561	5.638	5.709	5.774	5.833	5.889	5.941	5.990
40	5.163	5.216	5.266	5.313	5.358	5.439	5.513	5.581	5.642	5.700	5.753	5.803	5.849
60	5.056	5.107	5.154	5.199	5.241	5.319	5.389	5.453	5.512	5.566	5.617	5.664	5.708
120	4.950	4.998	5.044	5.086	5.126	5.200	5.266	5.327	5.382	5.434	5.481	5.525	5.568
inf	4.845	4.891	4.934	4.974	5.012	5.081	5.144	5.201	5.253	5.301	5.346	5.388	5.427

Respuestas a algunos ejercicios impares

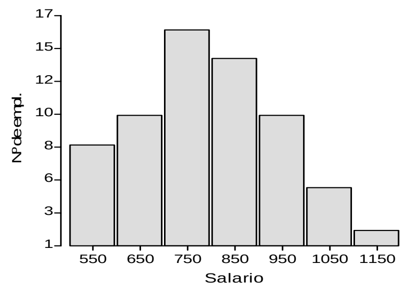
Capítulo 1

1.1: **Situación A**: 1) variables: número de pétalos de las flores de plantas de zapallo y producción de semillas por planta de zapallo; 2) variables: amarilleo de las hojas de plantas de zapallo y diámetro de los ovarios.

Situación B: variables: clorosis del follaje y altura de plántulas de tabaco.

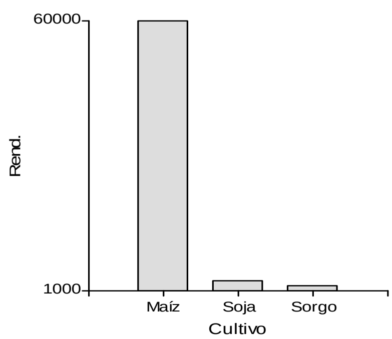
1.3: Variables discretas: casos a, e y f; variables continuas casos: b, c, d y g.

1.5: a)



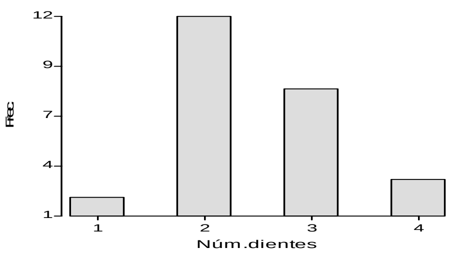
b) El 52.31%. No es el cuantil 0.80; c) El 47.69%; d) $X_{0.50} \cong \$750$; $X_{0.30} \cong \$700$.

1.7:



1.9: a) Tabla de Distribucion de Frecuencias

Variable		FA	FR	FAA	FRA
Núm.dientes	1	2	0.08	2	0.08
Núm.dientes	2	12	0.48	14	0.56
Núm.dientes	3	8	0.32	22	0.88
Núm.dientes	4	3	0.12	25	1.00



b) El 8%; c) El 44%.

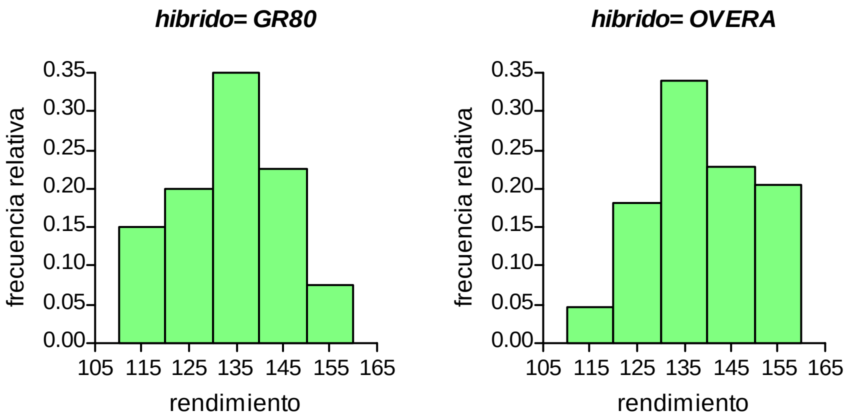
1.11:

a) Se recomendaría el híbrido Overa.

Estadística Descriptiva

Híbridos	Variable	n	Media	D.E.	Var	CV	Mín	Máx	Mediana
GR80	Rend.	40	133.90	11.44	130.81	8.54	110.00	152.00	136.00
Overa	Rend.	44	140.52	10.54	111.19	7.50	115.00	158.00	139.50

b) Para representar gráficamente las muestras se utilizaron histogramas de frecuencias relativas.



A los fines de facilitar la comparación se utilizaron los mismos intervalos de clase, comenzando en rendimientos de 110 y terminando en 160 y se igualaron los máximos de la escala de las frecuencias relativas. Se observa que la distribución de frecuencias de overá está desplazada hacia la derecha respecto de la de GR80, lo que se corresponde con su mayor valor promedio.

Capítulo 2

2.1:

a) $\Omega = \{ (x,y) / x = 1, 2, \ldots, 6; y = 1, 2, \ldots, 6 \}$

$=\{(1,1);(1,2);(1,3);(1,4);(1,5);(1,6);$

$(2,1);(2,2);(2,3);(2,4);(2,5);(2,6);$

$(3,1) ; \ldots ; (3,6);(4,1); \ldots ;(4,6);(5,1); \ldots ;(5,6);(6,1); \ldots ;(6,6)\}$

b) Es finito ya que $\#\Omega = 36$

c) Si. Es una variable aleatoria discreta.

2.3: $A= \{ (x,y) \in \Omega: x = 2$ ó $y = 2 \}$

$B = \{ (x,y) \in \Omega: (x + y) \leq 5 \}$

2.5: a) $\Omega = R^+$

b) Variable aleatoria continua.

2.7: a) Sí, ya que $A \cap B = \varnothing$; c) $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

2.9: Situación A: $P(A) \geq 0$ para todo evento A

Situación B: Si A, B, C y D son eventos mutuamente excluyentes y posibles en Ω, luego:

$P(A) + P(B) + P(C) + P(D) = 1$ pero en este caso no lo es ya que:

$$P(A) + P(B) + P(C) + P(D) = \frac{9}{120} + \frac{45}{120} + \frac{27}{120} + \frac{46}{120} > 1$$

2.11: a) Denotando 1 al éxito y 0 al fracaso del arranque se tiene que: $\Omega = \{ (0,0,0,0); (0,0,0,1); (0,0,1,0); (0,1,0,0); (1,0,0,0); (0,0,1,1); (0,1,1,0); (1,1,0,0); (0,1,0,1); (1,0,1,0); (1,0,0,1); (0,1,1,1); (1,1,1,0); (1,0,1,1); (1,1,0,1); (1,1,1,1) \}$

b) $P(\{w\}) = \dfrac{1}{16}$

c) $X: \Omega \rightarrow A \subseteq R$ donde $A = \{ 0, 1, 2, 3, 4 \}$

d) $P(X=3)=\frac{4}{16}=\frac{1}{4}$

$$P(X\leq 2)=P(X=0)+P(X=1)+P(X=2)=\frac{1}{16}+\frac{4}{16}+\frac{6}{16}=\frac{11}{16}$$

2.13: $P(A)=\frac{2000}{N}\Rightarrow 0.90=\frac{2000}{N}\Rightarrow N=\frac{2000}{0.90}=2222$

Se deberán adquirir 2222 plántulas.

Capítulo 3

3.1: a) 0.9032 ; b) 1 ; c) 0.0968 ; d) 0.68268 ; e) 0.14988, f) 0

3.3: a) 0.3085 ; b) 0.383

3.5: a) x = 17.022 micrones ; b) el 75% de la distribución de la variable diámetro de un sedimento, comprende valores menores o iguales a 17 micrones.

3.7: a) 0.2266 ; b) 0.2902

3.9: La estrategia A produce un 52% de los frutos de la Categoría II y la B un 55%. Se elige la estrategia B.

3.11: a) 0.6554 ; b) 0.1357 ; c) 204 cajones con una ganancia de \$530; d) 145 cajones con una ganancia de \$370. No es beneficiosa su aplicación.

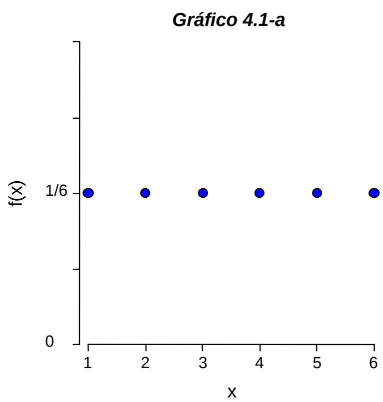
Capítulo 4

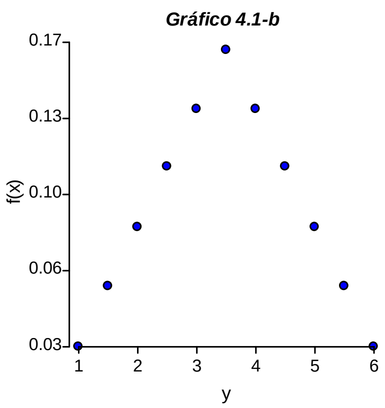
4.1:

a) Ω = {11, 12, 13, 14, 15, 16, 21, 22, 23, 24, 25, 26, 31, 32, 33, 34, 35, 36, 41, 42, 43, 44, 45, 46, 51, 52, 53, 54, 55, 56, 61, 62, 63, 64, 65, 66}

V.A.: $\overline{X}$	P($\overline{X}$=x)	V.A.: $\overline{X}$	P($\overline{X}$=x)	V.A.: $\overline{X}$	P($\overline{X}$=x)
1	1. $\frac{1}{36}$ = 0.027	3	5. $\frac{1}{36}$ = 0.138	5	3. $\frac{1}{36}$ = 0.083
1.5	2. $\frac{1}{36}$ = 0.055	3.5	6. $\frac{1}{36}$ = 0.166	5.5	2. $\frac{1}{36}$ = 0.055
2	3. $\frac{1}{36}$ = 0.083	4	5. $\frac{1}{36}$ = 0.138	6	1. $\frac{1}{36}$ = 0.027
2.5	4. $\frac{1}{36}$ = 0.111	4.5	4. $\frac{1}{36}$ = 0.111		

b)





En el gráfico 4.1-a, puede verse la distribución de la variable *x = número de puntos al arrojar un dado* y en el gráfico 4.1-b, la distribución de la variable *y= media del número de puntos al arrojar un par de dados*.

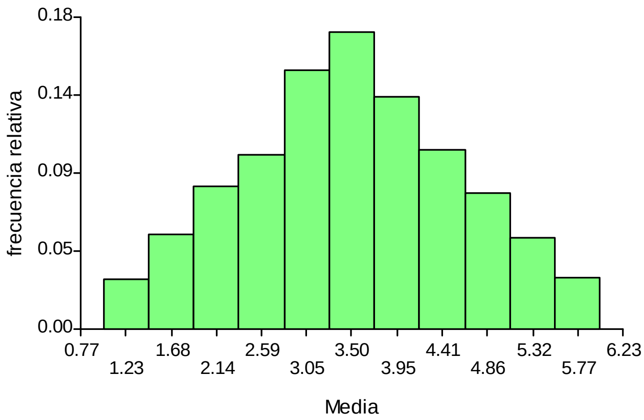
La esperanza de y=3.5 y varianza de y= 1.458.

Una forma alternativa de trabajar este ejercicio es mediante la simulación de un proceso de muestreo con muestras de tamaño n=2. Para ello utilizando InfoStat, construya un archivo de datos con una única columna con los números 1, 2, 3, 4, 5 y 6 que representan los valores posibles del número de puntos en la cara de un dado. Utilizando la opción del meno Aplicaciones>Didácticas>Remuestreo podremos simular el proceso de arrojar dos dados y calcular la media de sus puntos. Al activarse esta opción aparecerá una ventana que nos pide que especifiquemos la variable sobre la que queremos hacer el muestreo. Seleccionaremos de la lista que aparece a la izquierda del la pantalla el nombre de la columna que contiene los datos que ingresamos previamente y accionando el botón con el símbolo (→) la pasaremos en una lista que aparece a la derecha de la pantalla y que esta encabezada con el título *Muestra*. Una vez realizada esta operación accional el botón **Aceptar**. A continuación aparecerá una ventana de diálogo en la que seleccionaremos: Tipo de muestreo: *Aleatorio con reposición,* Guardar: *Media*, Nro muestras: *5000* y Tamaño muestra: *2*, como se ilustra en la siguiente figura.

Al aceptar aparecerá una nueva tabla de datos que contiene dos columnas, una encabezada con el rótulo *Muestra* y otra encabezada con el rótulo *Media*. La columna que nos interesa es esta última porque contiene las 5000 medias que obtuvimos por simulación del muestreo con reposición de tamaño 2 (simulamos 5000 veces que arrojábamos el dado dos veces y calculamos la media del puntaje obtenido en ambas oportunidades). La elección de que el muestreo sea *con reposición* permite que la siguiente vez que se arroja el dado pueda salir eventualmente el mismo número que la primera vez.

La media y varianza muestrales de estos 5000 datos, debería ser muy parecida a la esperanza y varianza teórica calculada anteriormente (3.5 y 1.458 respectivamente). Verifíquelo pidiendo una estadística descriptiva de la variable *Media.*

Si queremos ver la forma de la distribución de la media muestral, podemos seleccionar el menú Gráficos>Histograma e indicar que queremos un histograma de la variable *Media.* Obtendremos un gráfico como el siguiente (juegue con el número de intervalos en el que divide al eje X):



4.3: a) 0.0968 ; b) 7624 lts.

4.5: a) 0.005 ; b) 7300 lts.

4.7: a) 0.25 ; b) 0.09

4.9: 0.15

<u>*Capítulo 5*</u>

5.1: a) [58.45 ; 61.55] si α = 0.05; [57.96 ; 62.04] si α = 0.01; b) [59.02 ; 60.98] amplitud = 1.96 ; c) [57.83 ; 62.17] amplitud = 4.34

5.3: a) n $\cong$ 32 ; b) n $\cong$ 55

5.5: a) si $\bar{x}$ = 24.1 el intervalo será: [23.12 ; 25.08] amplitud = 1.96; si $\bar{x}$ = 25.5 el intervalo será: [24.52 ; 26.48] amplitud = 1.96; si $\bar{x}$ = 23 el intervalo será: [22.02 ; 23.98] amplitud = 1.96; si $\bar{x}$ = 24 el intervalo será: [23.02 ; 24.98] amplitud = 1.96; si $\bar{x}$ = 25.9 el intervalo será: [24.92 ; 26.88] amplitud = 1.96

b) si $\bar{\bar{x}}$ = 24.5 el intervalo será: [24.06 ; 24.96] amplitud = 0.88

c) La amplitud de las muestras individuales es mayor que la amplitud del intervalo de la muestra mayor.

5.7: a) n $\cong$ 18; b) n $\cong$ 71

<u>*Capítulo 6*</u>

6.1: a) Descartar H_0, Z=3.33; b) LI=17.06; LS=22.94; c) Se rechaza H_0;

d) LI=16.14,S=23.86; e) Se rechaza H_0. La media es mayor que 15.

6.3: a) Se aceptará incorrectamente la hipótesis nula 38 de cada 100 veces.

b) Aumentando el tamaño de la muestra.

6.5: a) H_0: $\mu = 45$ H_1: $\mu > 45$.

b) T= 4.86. $t_{19,0.99}$=2.539. Se rechaza H_0.

c) No se justifica realizar un cálculo de potencia ya que se rechazó H_0.

Capítulo 7

7.1: a) Unilateral derecha.

b) $\bar{X} = 221.63$

7.3: a) Para la prueba H_0: $\sigma^2 \geq 25$ vs. H_1: $\sigma^2 < 25$, con α= 0.10, el valor del estadístico calculado es 3.4, el punto crítico es 4.1682 por lo que se rechaza la hipótesis nula, lo que implica que la maquina cumple con esta especificación.

b) El intervalo unilateral derecho, para 1- α= 0.90, tiene LS = 20.35.

7.5: a) Al construir un intervalo unilateral derecho se obtiene LS = 0.0036, por lo que el suelo se considera homogéneo.

7.7: a) p=0.85

b) 0.4958

c) 6.6079

7.9: a) Para la prueba H_0: $\mu_{Nuevo} \leq \mu_{Control}$ vs. H_1: $\mu_{Nuevo} > \mu_{Control}$, con α= 0.10, el valor del estadístico calculado es 1.02. No se rechaza la hipótesis nula.

b) Muestras independientes provenientes de distribuciones normales con varianzas homogéneas.

c) El intervalo bilateral para $\mu_{Nuevo} - \mu_{Control}$, con 1- α= 0.90, tiene LI= -2.64 y LS= 9.44.

7.11: Para la prueba H_0: $\mu_{con\ poda} = \mu_{sin\ poda}$ vs. H_1: $\mu_{con\ poda} \neq \mu_{sin\ poda}$, con α=0.05, T=1.23. No hay evidencias suficientes para rechazar H_0, por lo tanto no hay efecto de la poda en el diámetro de los fustes.

7.13: El valor del estadístico es T=7.25. Se rechaza la hipótesis nula.

7.15: Para la prueba H_0: $\mu_{H1} = \mu_{H2}$ vs. H_1: $\mu_{H1} \neq \mu_{H2}$, con α= 0.05, T= -4.98. Se rechaza H_0.

Capítulo 8

8.1: a) H_0: $\tau_1 = \tau_2 = ... = \tau_6 = 0$ versus

H_1: al menos un tratamiento tiene efecto no nulo

b) **Cuadro de Análisis de la Varianza (SC tipo III)**

F.V.	SC	gl	CM	F	p-valor
Modelo	847.05	5	169.41	14.37	<0.0001
Inoculante	847.05	5	169.41	14.37	<0.0001
Error	282.93	24	11.79		
Total	1129.97	29			

Como el valor p asociado a la hipotesis de efecto nulo de inoculantes es <0.0001 se rechaza la hipótesis nula y se concluye, trabajando con un nivel de significación del 5%, que al menos un tratamiento tiene efecto no nulo. Se realiza la prueba de "a posteriori" de Fisher:

Test:LSD Fisher Alfa:=0.05 DMS:=4.48178

Error: 11.7887 gl: 24

Inoculante	Medias	n					
Cepa V	13.26	5	A				
Cepa III	14.64	5	A	B			
Cepa VI	18.70	5		B	C		
Cepa IV	19.92	5			C	D	
Cepa II	23.98	5				D	
Cepa I	28.82	5					E

Letras distintas indican diferencias significativas(p<= 0.05)

Como resultado de esta prueba se recomienda la Cepa I ya que prove cantidad de nitrógeno estadísticamente mayor que el resto de las cepas.

8.3: a) $Y_{ij} = \mu + \tau_i + \varepsilon_{ij}$ donde:

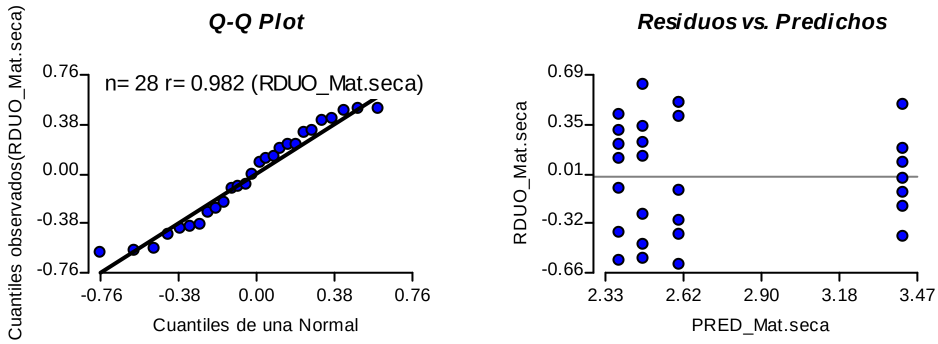
Y_{ij} = es la j-ésima observación de materia seca bajo la i-ésima carga animal.

μ= media general de materia seca.

τ_i = efecto de la i-ésima carga animal.

ε_{ij} = variable aleatoria normal, independientemente distribuida con esperanza cero y varianza σ^2 $\forall$ i j.

b) ε_{ij} están normal e independientemente distribuidos con esperanza cero y varianza σ^2. Para estudiar el cumplimiento de estos supuestos se recurre a métodos gráficos (QQ-plot para normalidad, Residuos vs predichos para homocedasticidad)



El análisis de las figuras precedentes perimitiría sustentar los supuestos normalidad y homogeneidad de varianzas.

c) **Cuadro de Análisis de la Varianza (SC tipo III)**

F.V.	SC	gl	CM	F	p-valor
Modelo	4.69	3	1.56	9.84	0.0002
Tratamiento	4.69	3	1.56	9.84	0.0002
Error	3.81	24	0.16		
Total	8.50	27			

Como p=0.0002 es menor que α=0.05 se rechaza la hipótesis de efectos de tratamientos nulos, es decir al menos un tratamiento produce efecto diferente. Se realiza la prueba “a posteriori” de Fisher:

Test: LSD Fisher Alfa:=0.05 DMS:=0.43964

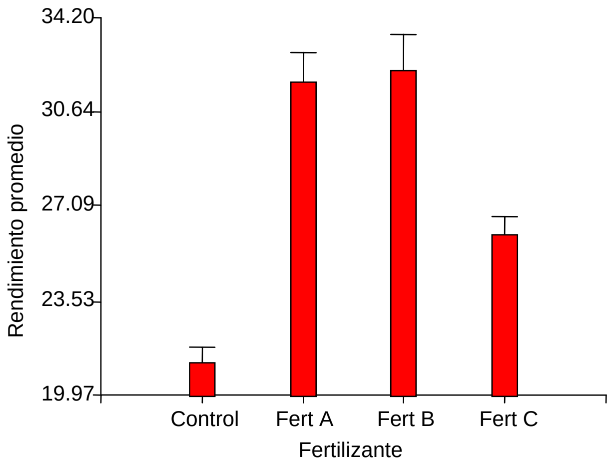
Error: 0.1588 gl: 24

Tratamiento	Medias	n		
carga8	2.39	7	A	
carga2	2.47	7	A	
carga6	2.60	7	A	
carga4	3.41	7		B

Letras distintas indican diferencias significativas(p<= 0.05)

Se recomienda la carga animal de 4 novillos/ha, porque es la carga que induce la mayor producción de materia seca, siendo estadísticamente diferente de la producción promedio inducida por resto de las cargas animales.

8.5: a)



Se puede observar que los fertilizantes A y B son los que producen mayores rendimientos medios.

b) **Cuadro de Análisis de la Varianza (SC tipo III)**

F.V.	SC	gl	CM	F	p-valor
Modelo	410.80	3	136.93	28.68	<0.0001
Fertilizante	410.80	3	136.93	28.68	<0.0001
Error	76.40	16	4.78		
Total	487.20	19			

Como el p-valor (<0.0001) es menor que α=0.05 se rechaza la hipótesis nula y se concluye que al menos un tratamiento tiene efecto no nulo. Se recomienda comparar las medias de tratamientos mediante una prueba de comparaciones múltiples. A continuación se presenta el resultado de la prueba de Fisher.

Test:LSD Fisher Alfa:=0.05 DMS:=2.92977

Error: 4.7750 gl: 16

Fertilizante	Medias	n			
Control	21.20	5	A		
Fert C	26.00	5		B	
Fert A	31.80	5			C
Fert B	32.20	5			C

Letras distintas indican diferencias significativas(p<= 0.05)

c) Los fertilizantes A y B son los que producen mayores rindes y sus medias de rendimiento no difieren estadísticamente entre sí. Luego, cualquiera de los dos puede ser recomendado.

8.7: CMDentro$=\dfrac{(n_1-1)S_1^2+...+(n_a-1)S_a^2}{(n_1-1)+...+(n_a-1)}=\dfrac{9(0.032)+...+9(0.020)}{54}=0.019$

CMEntre = $S_{\bar{x}}^{2}.n$ = 0.075x10 = 0.75;   Fobs.= 39.47 con 5 y 54 gl.

p-valor <0.0001.

Diferencia mínima significativa (Fisher)=0.123. La matriz de distancias absolutas entre medias es:

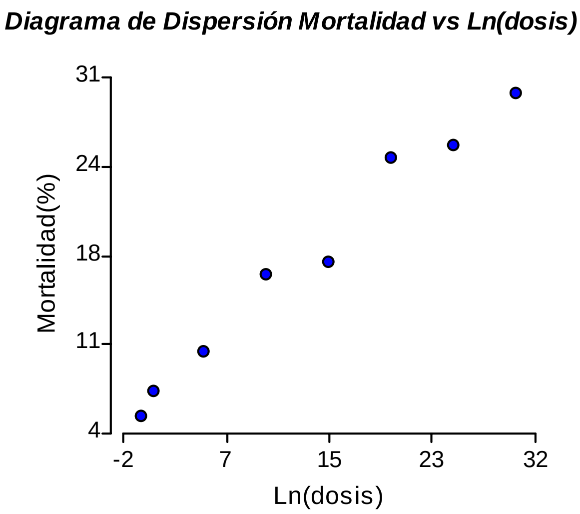
	1	3	5	6	10	20
1	0.00					
3	0.00	0.00				
5	0.23	0.23	0.00			
6	0.53	0.53	0.76	0.00		
10	1.17	1.17	0.94	1.70	0.00	
20	2.21	2.21	1.98	2.74	1.04	0.00

Luego, solo a densidades 1 y 3 los pesos promedios son indistinguibles. El resto de las densidades producen pesos promedios decrecientes.

<u>*Capítulo 9*</u>

9.1:

a)



Análisis de regresión lineal

Variable	N	R^2
Mortalidad(%)	8	0,98

Coeficientes de regresión y estadísticos asociados

Coef	Est.	EE	LI(95%)	LS(95%)	T	p-valor
const	6,05	0,85	3,98	8,12	7,15	0,0004
Ln(dosis)	0,83	0,05	0,70	0,95	16,47	<0,0001

Cuadro de Análisis de la Varianza

F.V.	SC	gl	CM	F	p-valor
Modelo	594,84	1	594,84	271,13	<0,0001
Ln(dosis)	594,84	1	594,84	271,13	<0,0001
Error	13,16	6	2,19		
Total	608,00	7			

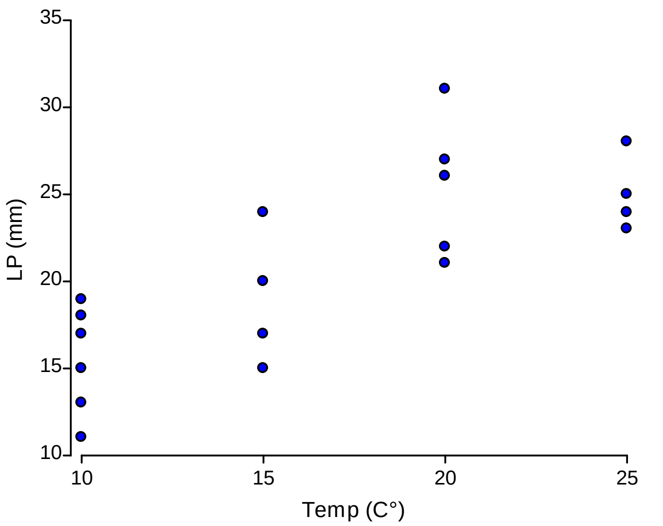
b) El diagrama de dispersión sugiere que es razonable proponer un modelo de regresión lineal. c) El modelo lineal es: Mortalidad $_i$ = α + β Ln(dosis)$_i$+ε_i con el supuesto de que los términos de error ε_i son variables aleatorias independientes con distribución normal de media cero y varianza σ^2. d) Los estimadores de los parámetros (coeficientes) del modelo son a=6,05 (estimador de α, ordenada al origen) y b=0,83 (estimador de β, pendiente).El estimador de σ^2 es el CME=2,19. e) Desde el cuadro de ANAVA se desprende que el Modelo explica una parte significativa de la variabilidad de los datos El p-valor (<0,0001) asociado a la hipótesis nula, que postula que las variaciones en mortalidad no son explicadas por la relación lineal con la dosis, induce al rechazo de esta hipótesis. Realizar esta observación es equivalente a concluir que la pendiente de la relación lineal es estadísticamente distinta de cero.

9.3:

a) En el experimento del ejemplo anterior se registra un solo valor de Y para cada X, en este ejemplo se tomaron varios valores de Y (longitud plántula) para cada valor de X (temperatura). Luego este conjunto de datos también podría analizarse con ANAVA para un modelo de efectos de tratamientos (temperatura)

b) El diagrama de dispersión sugiere que existe una tendencia lineal de la longitud de plántulas en el rango de temperaturas usadas en el experimento.

Diagrama de Dispersión de Longitud Plántula vs. Temperatura



c) El modelo lineal es: $LP_{ij} = \alpha + \beta Temperatura_i + \varepsilon_{ij}$ con el supuesto de que los términos de error ε_{ij} son variables aleatorias independientes con distribución normal de media cero y varianza σ^2. Los estimadores de los parámetros (coeficientes) del modelo son a=8,69 (estimador de α, ordenada al origen) y b=0,72 (estimador de β, pendiente).

Análisis de regresión lineal

Variable	N	R²
LP (mm)	19	0,60

Coeficientes de regresión y estadísticos asociados

Coef	Est.	EE	LI(95%)	LS(95%)	T	p-valor
const	8,69	2,54	3,32	14,06	3,42	0,0033
Temp (C°)	0,72	0,14	0,42	1,02	5,04	0,0001

Cuadro de Análisis de la Varianza

F.V.	SC	gl	CM	F	p-valor
Modelo	317,86	1	317,86	25,41	0,0001
Temp (C°)	317,86	1	317,86	25,41	0,0001
Error	212,66	17	12,51		
Total	530,53	18			

d) Desde el cuadro de ANAVA se desprende que el Modelo explica una parte significativa de la variación en los datos, dado que el valor-p asociado a la hipótesis nula que postula que las variaciones en LP no son explicadas por la relación lineal con la temperatura, es menor que el nivel de significación propuesto. La recta ajustada expresa el valor esperado de LP para cada temperatura. Como tiene pendiente positiva, a mayor temperatura se debe esperar mayor longitud, i.e. a 25°C deberíamos esperar que las plantas germinadas muestren mayor vigor.

9.5:

a) El insecticida es igualmente efectivo ya que, en ambas especies, mueren la misma cantidad de insectos por cada ppm de insecticida que se aplica.

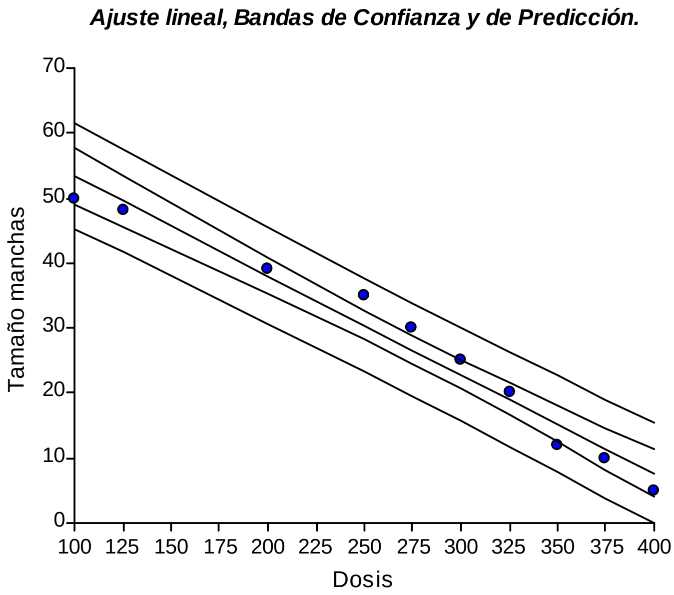
b) La sobrevida inicial (cuando no se agrega insecticida) es diferente en cada especie. Podríamos pensar que la especie B tiene una menor sobrevida natural. Ambas especies son igualmente afectadas frente un aumento de concentración en el rango de concentraciones ensayadas en el experimento. La respuesta es proporcional a la concentración del insecticida.

c) La sobrevida disminuye en 15 unidades por cada ppm en la que se incrementa la concentración.

b) Con 4 ppm, se espera que: la Especie B no sobreviva y que la Especie A tenga una sobrevida de 20.

9.7:

a) El diagrama de dispersión sugiere que existe una tendencia lineal de pendiente negativa que modela el tamaño de las manchas en función de la dosis de fungicida usada en el experimento (mayor dosis, menor tamaño de mancha). Los estimadores de los parámetros (coeficientes) del modelo son a=68,49 (estimador de α, ordenada al origen) y b=-0,15 (estimador de β, pendiente). Desde el cuadro de ANAVA se desprende que el Modelo explica una parte significativa de la variación en el tamaño de las manchas ($P<0,0001$). En la siguiente figura, se presenta el ajuste (recta central), las bandas de confianza (alrededor de la recta de ajuste) y las bandas de predicción (bandas exteriores).



b) Desde la recta ajustada se predice que el tamaño de la mancha para 260 gr.p.a/ha seria Y=68,49-0,15*260=29,49.

Análisis de regresión lineal

Variable	N	R^2
Daño	10	0,97

Coeficientes de regresión y estadísticos asociados

Coef	Est.	EE	LI(95%)	LS(95%)	T	p-valor
const	68,49	2,79	62,06	74,92	24,56	<0,0001
Dosis	-0,15	0,01	-0,17	-0,13	-15,65	<0,0001

Cuadro de Análisis de la Varianza

F.V.	SC	gl	CM	F	p-valor
Modelo	2165,70	1	2165,70	245,06	<0,0001
Dosis	2165,70	1	2165,70	245,06	<0,0001
Error	70,70	8	8,84		
Total	2236,40	9			

Capítulo 10

10.1: a) **Estadística descriptiva**

Suplem.	Variable	n	Media	D.E.	Mín	Máx	Mediana
A1	PBruta	6	3.16	0.12	3.03	3.30	3.15
A2	PBruta	6	3.15	0.15	2.93	3.33	3.16
B1	PBruta	6	3.34	0.09	3.22	3.45	3.32
B2	PBruta	6	3.38	0.12	3.20	3.54	3.37
Control	PBruta	6	3.24	0.13	3.10	3.48	3.22

b) $Y_{ij} = \mu + \tau_i + \beta_j + \varepsilon_{ij}$ donde:

Y_{ij} = es la observación de proteína bruta bajo el i-ésimo suplemento, j-ésimo bloque (tambo).

μ= media general de proteína bruta.

τ_i = efecto del i-ésimo suplemento.

β_j = efecto del j-ésimo bloque (tambo).

ε_{ij} = variable aleatoria normal, independientemente distribuida con esperanza cero y varianza σ^2 $\forall$ i j. El modelo supone que los tratamientos no interactuan con los bloques.

Análisis de la varianza

Variable	N	R²	R² Aj	CV
PBruta	30	0.86	0.80	2.05

Cuadro de Análisis de la Varianza (SC tipo III)

F.V.	SC	gl	CM	F	p-valor
Modelo	0.55	9	0.06	13.76	<0.0001
Suplementos	0.26	4	0.07	14.72	<0.0001
Tambo	0.29	5	0.06	13.00	<0.0001
Error	0.09	20	4.4E-03		
Total	0.64	29			

Como $p<0.0001$ es menor que $\alpha=0.05$ se rechaza la hipótesis de efectos nulos y se concluye que al menos un suplemento tiene efecto no nulo. Se realiza la prueba "a posteriori":

Test:LSD Fisher Alfa:=0.05 DMS:=0.08009

Error: 0.0044 gl: 20

suplementos	Medias	n			
A2	3.15	6	A		
A1	3.16	6	A		
Control	3.24	6		B	
B1	3.34	6			C
B2	3.38	6			C

Letras distintas indican diferencias significativas($p<= 0.05$)

Los suplementos B1 y B2 son los que producen mayores contenidos de proteína bruta y no difieren estadísticamente. Luego, se recomienda cualquiera de los dos.

10.3:

a) $Y_{ijk}=\mu+\alpha_i+\beta_j+\delta_{ij}+\varepsilon_{ijk}$ con i=1,2; j=1,…4; k=1,..,4

donde:

Y_{ijk} representa el largo de la cola del nemátodo de la k-ésima repetición, en el i-ésimo nivel del factor Sexo y j-ésimo nivel de factor Temperatura

μ representa una media general del largo de la cola

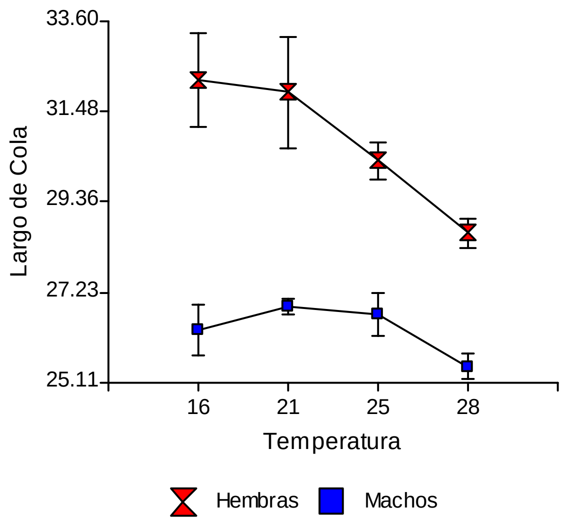
α_i es el efecto que produce el i-ésimo nivel del factor Sexo

β_j corresponde al efecto del j-ésimo nivel del factor Temperatura

δ_{ij} los efectos adicionales (interacciones) para cada combinación de los niveles de los factores

ε_{ijk} es el error asociado a la observación ijk-ésima que se supone normal e independiente con esperanza cero y varianza común σ^2

b)



c) **Análisis de la varianza**

Variable	N	R^2	R^2 Aj	CV
Largo de Cola	32	0.80	0.74	4.97

Cuadro de Análisis de la Varianza (SC tipo III)

F.V.	SC	gl	CM	F	p-valor
Modelo	192.98	7	27.57	13.68	<0.0001
Temperatura	27.99	3	9.33	4.63	0.0108
Sexo	155.32	1	155.32	77.07	<0.0001
Temperatura*Sexo	9.66	3	3.22	1.60	0.2159
Error	48.37	24	2.02		
Total	241.34	31			

Primeramente se analiza la interacción, en este caso p=0.2159 es mayor α=0.05, con lo cual se concluye que no hay interacción entre los factores Sexo y Temperatura.
Se observan luego los efectos principales, el factor sexo es significativo (p<0.0001) y el factor Temperatura también (p=0.0108). Se realiza entonces una prueba "a posteriori" para este factor ya que el otro tiene solo dos niveles:

Test:LSD Fisher Alfa:=0.05 DMS:=1.46497
Error: 2.0153 gl: 24

Temperatura	Medias	n		
28	27.06	8	A	
25	28.53	8	A	B
16	29.29	8		B
21	29.41	8		B

Letras distintas indican diferencias significativas(p<= 0.05)

Se concluye que hay diferencias entre sexo y que bajo la temperatura de 28 ºC los nematodes preserntan un largo una cola significativamente menor que bajo temperatures de 16 y 21 ºC.

10.5:
Para especificar el modelo del diseño en parcelas divididas en bloques de este ejercicio en el paquete estadístico infostat, se debería escribir las siguientes instrucciones:

Bloque

Fecha\Bloque*Fecha

Bloque*Fecha

Variedad

Fecha*Variedad

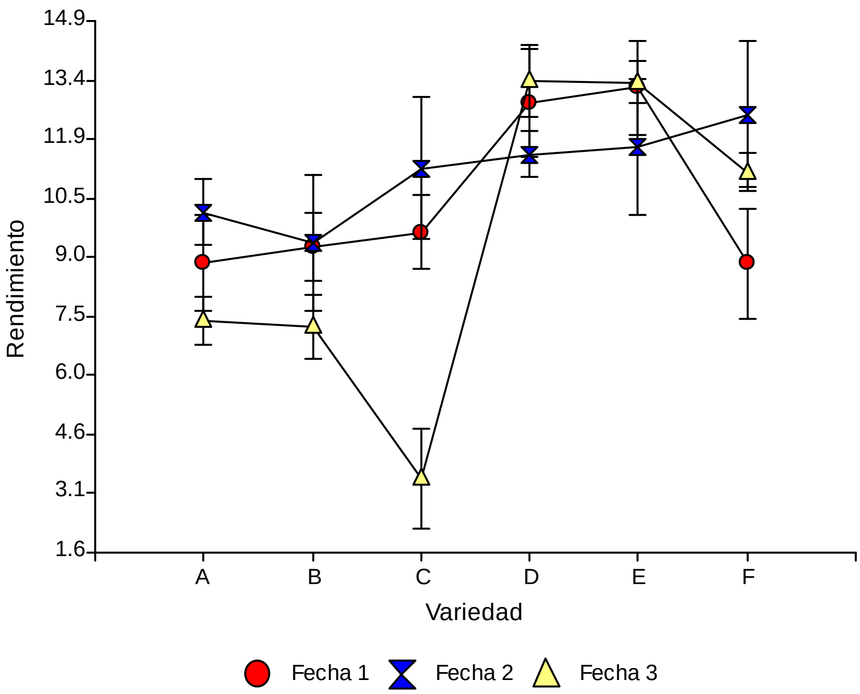
La tabla de análisis de la varianza resultante se presenta a continuación:

Variable	N	R^2	R^2 Aj	CV
Rend	72	0.70	0.53	21.87

Cuadro de Análisis de la Varianza (SC tipo III)

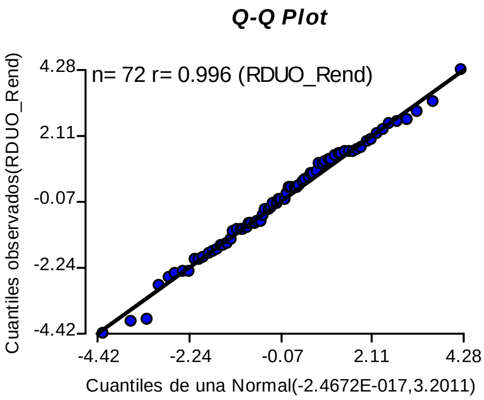
F.V.	SC	gl	CM	F	p-valor	(Error)
Modelo	535.12	26	20.58	4.08	<0.0001	
Bloque	29.34	3	9.78	1.94	0.1372	
Fecha	38.00	2	19.00	2.62	0.1524	(Bloque*Fecha)
Bloque*Fecha	43.57	6	7.26	1.44	0.2216	
Variedad	260.51	5	52.10	10.32	<0.0001	
Fecha*Variedad	163.70	10	16.37	3.24	0.0032	
Error	227.28	45	5.05			
Total	762.40	71				

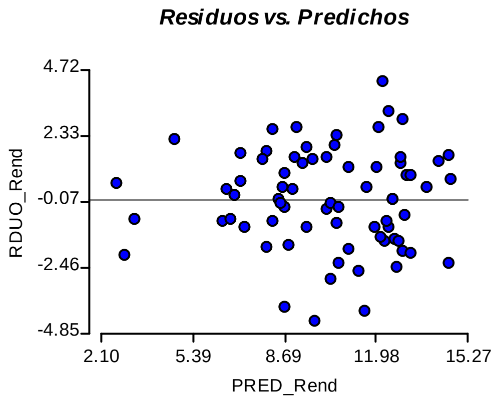
La interacción entre Fecha y Variedad fue significativa (p=0.0032), esto indica que los factores intervinientes no actúan independientemente. Es por ello que no se establecerán conclusiones sobre los efectos principales. Se examinarán los niveles del factor Fecha manteniendo fijos los niveles del factor Variedad. La siguiente figura muestra los valores medios de las combinaciones de niveles de ambos factores:



La figura permite visualizar la interacción Fecha x Variedad. Las variedades que se muestran mejores, al menos para las Fecha 1 y 3 son las variedades D y E. Para la Fecha 1, las mejores variedades son F, E D y C. Un análisis más detallado requiere el examen de comparaciones múltiples.

Las figuras subsiguientes permiten analizar los supuestos. Se puede observar que no habría inconvenientes en el cumplimiento tanto de la normalidad como de la homogeneidad de varianzas.





Capítulo 11

11.1: $\chi 2=26.90$; $\nu=2$

11.3: $\chi 2=10.39$; $\nu=2$

11.5: $\chi 2=1.467$; $\nu=3$

Esta obra terminó de imprimir en el mes de
Marzo de 2005 en Editorial Brujas.
Córdoba, Argentina

View publication stats