IN 540
Métodos Estadísticos
para Economía y Gestión

Análisis de Varianza (ANOVA)

Ref: Cáp. X, “Estadística aplicada los negocios y la economía”, Allen L. Webster (Irwin MacGraw-Hil 2000)

10.1 Introducción

ANOVA está diseñada específicamente para probar si dos o más poblaciones tienen la misma media. Aun cuando el propósito de ANOVA es hacer pruebas para hallar las diferencias en las medias poblacionales, implica un examen de las varianzas muestrales; de allí el término *análisis de varianza*. Más específicamente, el procedimiento se puede utilizar para determinar si cuando se aplica un "tratamiento" en particular a una población, éste tendrá un impacto significativo en su media. El uso de ANOVA originado en el campo de la agricultura, en donde el término *tratamiento* se utiliza de la misma manera que cuando se tratan varias parcelas de tierra con diferentes fertilizantes y se notan las diferencias en los rendimientos promedio de los cultivos. Hoy en día el término *tratamiento* se utiliza ampliamente, para referirse al tratamiento de los clientes respecto a diferentes despliegues publicitarios observando las diferencias subsiguientes en las compras promedio, o también al tratamiento de tres grupos de empleados ante tres tipos diferentes de programas de capacitación y observando las diferencias que ocurren en los niveles promedio de productividad o, en general, en toda situación en la cual se desea una comparación de las medias.

Consideremos como ejemplo el interés en medir los efectos relativos en la producción de los empleados de tres programas de capacitación. Estos tres tipos de formación adicional pueden ser 1) autodidactas, 2) impartido por computador, o 3) enseñado por un supervisor. En un estudio ANOVA, las **unidades experimentales** son los objetos que reciben el *tratamiento*. En nuestro ejemplo sobre capacitación, los empleados constituyen las unidades experimentales. El **factor** es la fuerza o variable cuyo impacto en tales unidades experimentales se desea medir. En este caso, "capacitación" es el factor de interés. Finalmente, los tres tipos de capacitación constituyen los **tratamientos**, o niveles del factor, del factor "capacitación".

La forma como se seleccionan los tratamientos determina si se está utilizando un **modelo de efectos fijos** o un **modelo de efectos aleatorios**. El modelo descrito anteriormente sobre el programa de capacitación para los empleados es un modelo de efectos fijos. Los tres programas de entrenamiento se seleccionaron o se fijaron antes de realizar el estudio. Se sabe cuál de los tres programas se desea probar desde el comienzo del estudio. Las conclusiones del estudio se aplican sólo a los tres programas incluidos en el estudio.

Modelo de efectos fijos En el cual se seleccionan tratamientos específicos o se fijan antes del estudio.

En contraste, se supone que Apex Manufacturing tenía muchos programas de capacitación distintos, y deseaban saber si los programas de capacitación en general tenían efectos diferentes en el desempeño de los empleados. Estos tres programas de capacitación utilizados en el estudio se consideran como una muestra de todos los programas de entrenamiento que la firma puede utilizar. No importa cuál de los tres métodos se utilice en el estudio para hacer la comparación. Toda conclusión del estudio se considera aplicable a toda la población de programas de capacitación. Este procedimiento produciría un modelo de efectos aleatorios.

Modelo de efectos aleatorios En el cual los niveles (*tratamientos*) utilizados en el estudio se seleccionan aleatoriamente de una población de niveles posibles.

Para la aplicación de ANOVA son esenciales tres suposiciones:

1. Todas las poblaciones involucradas son normales.
2. Todas las poblaciones tienen la misma varianza.
3. Las muestras se seleccionan independientemente.

Si un número de tratamientos se designa como c, el conjunto de hipótesis de prueba es

$$H_0\text{: } \mu_1 = \mu_2 = \mu_3 \cdots = \mu_c$$
$$H_A\text{: No todas las medias son iguales}$$

La letra c se utiliza para el número de tratamientos debido a que en una tabla ANOVA, que se verá en breve, cada tratamiento se especifica en su propia columna.

Se puede argumentar que sería posible probar la igualdad de varias medias utilizando varias pruebas t con dos muestras, tal y como se hizo en el capítulo 9. Sin embargo, algunas complicaciones pueden hacer que este método no sea efectivo. Por ejemplo, si un fabricante desea comparar la producción diaria promedio para tres plantas, puede probar los tres siguientes conjuntos de hipótesis:

$$H_0\text{: } \mu_1 = \mu_2$$
$$H_A\text{: } \mu_1 \neq \mu_2$$

y

$$H_0\text{: } \mu_1 = \mu_3$$
$$H_A\text{: } \mu_1 \neq \mu_3$$

y

$$H_0\text{: } \mu_2 = \mu_3$$
$$H_A\text{: } \mu_2 \neq \mu_3$$

Si la hipótesis nula no se rechaza en cada una de las pruebas, se puede concluir que las tres medias son iguales.

Por lo menos dos problemas surgen con este método. Primero, si el número de poblaciones (plantas) se incrementa, el número de pruebas requeridas sube notablemente. Si hay cuatro plantas que el fabricante desea comparar, el número de pruebas individuales se duplica de 3 a ${}_4C_2 = 6$ pruebas. Surge el segundo problema y quizá el más molesto. Debido a un compuesto del valor α, el cual es la probabilidad de un error tipo I. Si se van a realizar estas pruebas a un nivel del 5%, y hay tres poblaciones que requieren tres pruebas de hipótesis separadas, la probabilidad de un error tipo I supera el 5%. Puede calcularse como

$$\begin{aligned} P(\text{Tipo I}) &= [1 - (1 - 0.05)(1 - 0.05)(1 - 0.05)] \\ &= 1 - (0.95)^3 \\ &= 0.1426 \end{aligned}$$

Mientras que se desea probar a un nivel del 5%, la necesidad de hacer tres pruebas incrementó la probabilidad del error tipo I más allá de los límites aceptables.

10.2 Análisis de varianza a una vía: Diseño completamente aleatorizado

Hay varias formas en las cuales puede diseñarse un experimento ANOVA. Quizá el más común es el *diseño completamente aleatorizado* o ANOVA *a una vía*. El término proviene del hecho que varios sujetos o unidades experimentales se asignan aleatoriamente a diferentes niveles de un solo factor. Por ejemplo, varios empleados (unidades experimentales) pueden seleccionarse aleatoriamente para participar en diversos tipos (niveles diferentes) de un programa de capacitación (el factor).

El director administrativo de una gran empresa industrial desea determinar si los tres programas de capacitación distintos tienen efectos diferentes en los niveles de productividad de los empleados. Estos programas son los tratamientos que puede evaluar el análisis de varianza. Se seleccionan aleatoriamente 14 empleados y se asignan a uno de los tres programas. Al terminar la capacitación, cada empleado responde un examen para determinar su competencia. Se colocan cuatro empleados en el primer programa de capacitación, y cinco en cada uno de los otros dos programas. Cada uno de estos tres grupos se trata de manera independiente como muestras separadas. Los puntajes de la prueba aparecen en la tabla 10.1, junto con unos cuantos cálculos básicos.

De las 15 *celdas* en la tabla, 14 tienen entradas. La última celda del primer tratamiento es una celda vacía. Una celda identificada como X_{ij} en donde i es la fila y j es la columna en la cual se encuentra ubicada la celda. X_{32} es la entrada de la tercera fila y la segunda columna. Se ve que es 81. X_{51} es la celda vacía. El número de filas en cada columna se indica con una r y el número de columnas o tratamientos se indica con una c. En el caso actual, $r = 5$ y $c = 3$.

Tabla 10.1
Prueba del puntaje de los empleados

	Tratamientos		
	Programa 1	Programa 2	Programa 3
	85	80	82
	72	84	80
	83	81	85
	80	78	90
	**	82	88
Columna medias $\bar{X}_j$	$\bar{X}_1 = 80$	$\bar{X}_2 = 81$	$\bar{X}_3 = 85$

Como se observa en la tabla 10.1, la media se calcula para cada tratamiento (columna). Debido a que las columnas se identifican mediante el subíndice j, el promedio de columnas se representa como $\bar{X}_j$. Finalmente, la gran media $\bar{\bar{X}}$ se calcula para todas las n observaciones.

La gran media de todas las observaciones del experimento

$$\bar{\bar{X}} = \frac{\Sigma X_{ij}}{n} \quad [10.1]$$

$$\bar{\bar{X}} = \frac{85 + 72 + 83 + \cdots + 90 + 88}{14}$$

$$= 82.14$$

El análisis de varianza se basa en una comparación de la cantidad de variación en cada uno de los tratamientos. Si de un tratamiento al otro la variación es significativamente alta, puede concluirse que los tratamientos tienen efectos diferentes en las poblaciones. En la tabla 10.1 se pueden identificar tres tipos o fuentes de variación. Vale la pena destacar que la primera variación es la suma de las otras dos.

1. Existe variación entre el número total de las 14 observaciones. No todos los 14 empleados tuvieron el mismo puntaje en la prueba. Esta se llama **variación total**.
2. Existe variación entre los diferentes tratamientos (muestras). Los empleados del programa 1 no tuvieron el mismo puntaje que los del programa 2 y 3. Esto se denomina **variación entre muestras**.
3. Existe variación dentro de un tratamiento dado (muestra). No todos los empleados de la primera muestra tuvieron el mismo puntaje. Esto es lo que se denomina **variación dentro de la muestra**.

Es al comparar estas fuentes diferentes de variación que se puede utilizar el análisis de varianza para probar la igualdad de las medias de poblaciones diversas. Toda diferencia que los tratamientos puedan tener en la productividad de los empleados se detectará mediante una comparación de estas formas de variación.

A. Fundamentos del ANOVA

Para determinar si tratamientos diferentes tienen efectos diferentes en sus respectivas poblaciones, se hizo una comparación entre la variación dentro de las muestras (W/S) y la variación entre muestras (B/S). La variación en los puntajes dentro de una muestra dada puede ser producida por una variedad de factores: la habilidad innata de los empleados en dicha muestra, la motivación personal, los esfuerzos individuales y la destreza, el factor suerte, y una gran cantidad de otras circunstancias aleatorias. El tratamiento en sí mismo no producirá ninguna variación en las observaciones dentro de alguna muestra, debido a que todas las observaciones en dicha muestra reciben el mismo tratamiento.

Es un asunto diferente con la variación entre muestras. La variación en los puntajes entre muestras (de una muestra a la siguiente) puede producirse por el mismo factor aleatorio que la variación dentro de una muestra (motivación, destreza, suerte, etc.), más toda influencia adicional que puedan tener los tratamientos diferentes. Puede existir un **efecto del tratamiento** entre muestras debido a que cada muestra tiene un tratamiento diferente.

Efecto del tratamiento Como las muestras diferentes tienen tratamientos distintos, la variación entre las muestras puede ser producida por los efectos de tratamientos diferentes.

Si un efecto del tratamiento existe, puede detectarse comparando la variación entre las muestras y la variación dentro de las muestras. Si la variación entre las muestras es significativamente mayor que la variación dentro de las muestras, un fuerte efecto de tratamiento está presente. Esta diferencia entre la variación *entre* muestras y la variación *dentro* de las muestras es precisamente lo que mide el análisis de varianza. El análisis de varianza es una relación de la variación entre muestras con la variación dentro de las muestras. Si los tratamientos diferentes tienen efectos diferentes, la variación entre muestras crecerá, haciendo que la razón aumente. Esta razón se basa en la razón F presentada en la sección anterior.

La razón F tal y como se utiliza en ANOVA La razón F es una razón de la variación entre muestras y la variación dentro de las muestras.

De nuevo, la variación entre muestras puede ser producida en parte por tratamientos diferentes. La variación dentro de una muestra dada puede ser producida sólo por factores aleatorios como la suerte, la destreza, y la motivación de los empleados. Dicha variación es independiente del tratamiento (ya que todas las observaciones dentro de una muestra tienen el mismo tratamiento) y es el resultado sólo del **error de muestreo aleatorizado** dentro de la muestra.

> **La razón F** Cuando las medias poblacionales son diferentes, el efecto del tratamiento está presente y las desviaciones entre las muestras serán grandes comparadas con la desviación del error dentro de una muestra. Por tanto, el valor F aumentará, lo cual es una razón de la variación del tratamiento y de la variación del error.

La variación total es igual a la variación producida por los tratamientos diferentes, más la variación producida por elementos de error aleatorios dentro de los tratamientos, como la destreza, la suerte y la motivación. Es decir,

Variación total = variación del tratamiento + variación del error

B. La suma de cuadrados

El reconocimiento de estas tres fuentes de variación permite la *división de la suma de cuadrados*, un procedimiento que es necesario para el análisis de varianza. Cada uno de los tres tipos de variación produce una suma de cuadrados. Existe: 1) la suma de cuadrados total (SCT), 2) la suma de cuadrados de los tratamientos ($SCTR$), y 3) la suma de cuadrados del error (SCE). Como era de esperarse

$$SCT = SCTR + SCE$$

Esto ilustra que SCT puede dividirse en sus dos componentes: $SCTR$ y SCE.

Se pueden utilizar estas sumas de cuadrados para probar la igualdad de las medias poblacionales. Vale la pena recordar del capítulo 3 que la varianza muestral se calcula así:

Varianza muestral
$$s^2 = \frac{\Sigma(X_i - \overline{X})^2}{n - 1} \qquad [10.2]$$

El numerador es la suma de los cuadrados de las desviaciones de la media. De esta forma, la suma de los cuadrados se utiliza para medir la variación. El denominador es el número de grados de libertad. Esta ecuación sirve como patrón que puede aplicarse a la suma de cuadrados en análisis de varianza.

Sea X_{ij} la observación *i*ésima en la muestra *j*ésima. Por ejemplo , X_{21} es la segunda observación en la primera muestra. En la tabla 10.1, $X_{21} = 72$, $X_{32} = 81$, $X_{43} = 90$, y así sucesivamente. Entonces,

Suma de cuadrados total
$$SCT = \sum_{i=1}^{r}\sum_{j=1}^{c} (X_{ij} - \overline{\overline{X}})^2 \qquad [10.3]$$

La gran media se le resta a cada una de las 14 observaciones. Las diferencias se elevan al cuadrado y se suman. Como lo muestra el signo de doble sumatoria en la fórmula (10.3), esto se hace a través de las filas y a través de todas las columnas. De allí en adelante la notación para los signos de sumatoria se elimina, en aras de la simplicidad. Utilizando los datos de la tabla 10.1 se tiene que:

$$\begin{aligned} SCT &= (85 - 82.14)^2 + (72 - 82.14)^2 + (83 - 82.14)^2 \\ &\quad + (80 - 82.14)^2 + (80 - 82.14)^2 + (84 - 82.14)^2 \\ &\quad + \cdots (90 - 82.14)^2 + (88 - 82.14)^2 \\ &= 251.7 \end{aligned}$$

Debería notarse que *SCT* es simplemente la variación de las observaciones alrededor de la gran media.

Para la suma de cuadrados de los tratamientos se tiene que:

Suma de cuadrados de los tratamiento	$SCTR = \Sigma r_j(\bar{X}_j - \bar{\bar{X}})^2$	[10.4]

El número de observaciones o filas en cada tratamiento, r_j, se multiplica por las diferencias cuadradas entre la media de cada tratamiento, $\bar{X}_j$, y la gran media. Los resultados se suman para todos los tratamientos. La fórmula (10.4) pide que se multiplique el número de filas en la *j*ésima columna (vale la pena recordar que *j* denota una columna) por la desviación de la media elevada al cuadrado de dicha columna de la gran media. La tabla 10.1 da

$$\begin{aligned} SCTR &= 4(80 - 82.14)^2 + 5(81 - 82.14)^2 + 5(85 - 82.14)^2 \\ &= 65.7 \end{aligned}$$

SCTR refleja la variación en las medias de la columna alrededor de la gran media.

La suma de cuadrados del error se expresa como

Suma del cuadrado del error	$SCE = \Sigma\Sigma(X_{ij} - \bar{X}_j)^2$	[10.5]

La media de un tratamiento, $\bar{X}_j$, se resta de cada observación en dicho tratamiento. Las diferencias se elevan al cuadrado y se suman. Esto se hace para todos los tratamientos, y los resultados se suman. Utilizando los datos de la tabla 10.1 nuevamente, se tiene que

$$SCE = (85 - 80)^2 + (72 - 80)^2 + (83 - 80)^2 + (80 - 80)^2$$

Para el primer tratamiento

$$+ (80 - 81)^2 + (84 - 81)^2 + (81 - 81)^2 + (78 - 81)^2 + (82 - 81)^2$$

Para el segundo tratamiento

$$+ (82 - 85)^2 + (80 - 85)^2 + (85 - 85)^2 + (90 - 85)^2 + (88 - 85)^2$$

Para el tercer tratamiento

$$= 186.0$$

SCE mide la variación aleatoria de los valores dentro de un tratamiento alrededor de su propia media.

Una revisión rápida de todos estos cálculos puede hacerse como

$$SCT = SCTR + SCE$$

$$251.7 = 65.7 + 186.0$$

Si se confía en la aritmética, se puede encontrar que SCE es simplemente

$$SCE = SCT - SCTR = 251.7 - 65.7 = 186.0$$

C. Cuadrados medios

Como lo dice la fórmula (10.2) para la varianza, después de obtener la suma de los cuadrados, cada una se divide por sus grados de libertad. Una suma de cuadrados dividida por sus grados de libertad produce un cuadrado medio. Es decir, si se divide una suma de cuadrados por sus grados de libertad, se obtiene un cuadrado medio.

Vale la pena recordar del capítulo 7 que se definió grados de libertad como el número total de observaciones del conjunto de datos menos toda "restricción" que pueda aplicarse. Una restricción fue todo valor que se calcula a partir del conjunto de datos.

En este aspecto, se nota que al calcular *SCT*, se utilizó todo el conjunto de datos de n observaciones para calcular un valor. Ese valor único era la gran media $\bar{\bar{X}}$, la cual representa una restricción. Por tanto *SCT* tiene $n-1$ grados de libertad.

El cálculo de *SCTR* involucra el uso de $c = 3$ medias muestrales de las cuales se puede calcular la gran media. Las medias muestrales por tanto se ven como puntos de datos individuales y la gran media se toma como restricción. *SCTR* tiene entonces $c-1$ grados de libertad.

Finalmente, se calculó *SCE* anteriormente sumando la desviación de $n = 14$ observaciones de $c = 3$ medias muestrales. Por tanto, *SCE* tiene $n-c$ grados de libertad.

Se nota que

$$\text{g.l. para } SCT = \text{g.l. para } SCTR + \text{g.l. para } SCE$$
$$n-1 = c-1+n-c$$

Como se anotó anteriormente, debido a que la suma de cuadrados dividida por sus grados de libertad produce un cuadrado medio, se halla la media total de los cuadrados, o cuadrado medio total, *CMT*

Cuadrado medio total	$CMT = \dfrac{SCT}{n-1}$	[10.6]

el cuadrado medio del tratamiento (*CMTR*) es

Cuadrado medio del tratamiento	$CMTR = \dfrac{SCTR}{c-1}$	[10.7]

y el cuadrado medio del error (*CME*) es

Cuadrado medio del error	$CME = \dfrac{SCE}{n-c}$	[10.8]

Utilizando los datos de la tabla 10.1 se tiene que:

$$\begin{aligned} CMT &= \frac{SCT}{n-1} \\ &= \frac{251.7}{14-1} \\ &= 19.4 \end{aligned}$$

$$\begin{aligned} CMTR &= \frac{SCTR}{c-1} \\ &= \frac{65.7}{3-1} \\ &= 32.9 \end{aligned}$$

$$CME = \frac{SCE}{n - c}$$
$$= \frac{186.0}{14 - 3}$$
$$= 16.9$$

Estos tres cuadrados medios están modelados a partir de la fórmula (10.2). Son sumas de los cuadrados divididas por sus grados de libertad, y como tales son varianzas. Es la razón de las dos últimas *CMTR* y *CME*, que se utiliza como base del análisis de varianza para probar la hipótesis respecto a la igualdad de las medias. Como se observó anteriormente, esta razón se ajusta a la distribución *F*, y se expresa como

Razón *F* para una prueba de medias	$F = \frac{CMTR}{CME}$	[10.9]

En el caso actual se vuelve

$$F = \frac{32.9}{16.9}$$
$$= 1.94$$

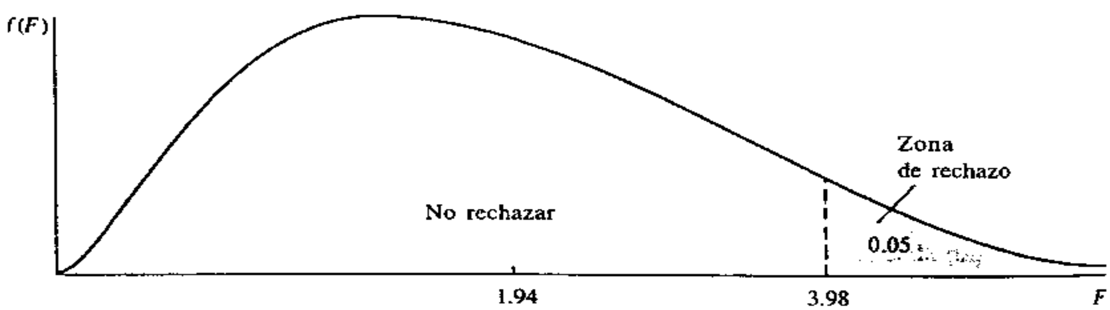
CMTR mide la variación entre tratamientos. Si los tratamientos tienen efectos diferentes, CMTR lo reflejará a través de su incremento. Entonces, la razón *F* en sí misma se incrementará. Por tanto, si la razón *F* se vuelve "significativamente" grande porque *CMTR* excede a *CME* por una cantidad grande, se reconoce que los efectos del tratamiento probablemente existen. Es probable que tratamientos diferentes tengan efectos diferentes en las medias de sus poblaciones respectivas, y podría rechazarse la hipótesis nula $\mu_1 = \mu_2 = \mu_3$.

El valor crítico de *F* que es considerado significativamente grande puede encontrarse en la tabla G (apéndice III) igual que antes. Se asume que el CEO desea probar las siguientes hipótesis a un nivel del 5%:

$$H_0\text{: } \mu_1 = \mu_2 = \mu_3$$
$$H_A\text{: No todas las medias son iguales}$$

Debido a que *CMTR* tiene $c - 1 = 3 - 1 = 2$ grados de libertad y *CME* tiene $n - c = 14 - 3 = 11$ grados de libertad, el valor crítico de F que se obtiene de la tabla es $F_{0.05,2,11} = 3.98$. El 2 se enumera antes del 11 al establecer los grados de libertad porque *CMTR* está en el numerador.

Figura 10.1
Los efectos de la capacitación



La regla de decisión representada en la figura 10.1 es

Regla de decisión: "No rechazar si $F \leq 3.98$. Rechazar la hipótesis nula si $F > 3.98$".

Debido a que se calculó que el valor *F* es de $1.94 < 3.98$, el CEO no debería rechazar la hipótesis nula. No puede rechazar a un nivel del 5% la hipótesis de que los puntajes de prueba promedio son los mismos para todos los tres programas de capacitación. No existe efecto significativo del tratamiento relacionado con alguno de los programas.

D. Una tabla de análisis de varianza

Es habitual resumir los cálculos del análisis de varianza en una tabla. El formato general de la tabla de análisis de varianza aparece en la tabla 10.2 A), mientras que la tabla 10.2 B) contiene los valores específicos del ejemplo sobre el programa de capacitación.

Vale la pena destacar que se enumeran las fuentes relevantes de la variación, y el valor F de 1.94 se muestra en la columna del extremo derecho. El ejemplo 10.1 proporciona una ilustración más concisa del análisis de varianza.

Tabla 10.2
Una tabla ANOVA resume los cálculos del análisis de varianza

A. La tabla de análisis de varianza generalizada

Fuente de variación	Suma de cuadrados	Grados de libertad	Cuadrado medio	Valor F
Entre muestras (tratamiento)	$SCTR$	$c-1$	$SCTR/(c-1)$	$CMTR/CME$
Dentro de muestras (error)	SCE	$n-c$	$SCE/(n-c)$	
Variación total	SCT	$n-1$		

B. Tabla de ANOVA para los programas de entrenamiento de empleados

Fuente de variación	Suma de cuadrados	Grados de libertad	Cuadrado medio	Valor F
Entre muestras (tratamiento)	65.7	2	32.9	1.94
Dentro de muestras (error)	186.0	11	16.9	
Variación total	251.7	13		

H_0: $\mu 1 = \mu 2 = \mu 3$

H_A: No todas las medias son iguales

Regla de decisión: No rechazar si $F \leq 3.98$. Rechazar si $F > 3.98$.

Conclusión: Ya que $F = 1.94 < 3.98$, no se rechaza la hipótesis nula.

Ejemplo 10.1

Robert Shade es vicepresidente de mercadeo en First City Bank, en Atlanta. Los recientes esfuerzos promocionales para atraer nuevos depositantes incluyen algunos juegos y premios en cuatro sucursales del banco. Shade está convencido de que diferentes tipos de premios atraerían a diferentes grupos de ingreso. Las personas de un nivel de ingreso prefieren los regalos, mientras que los de otro grupo de ingreso pueden sentirse más atraídas por viajes gratuitos a sitios favoritos para pasar vacaciones. Shade decide utilizar el monto de los depósitos como una medida representativa del ingreso. El desea determinar si existe una diferencia en el nivel promedio de depósitos entre las cuatro sucursales. Si se halla alguna diferencia, Shade ofrecerá una diversidad de premios promocionales.

Solución

Aquí aparecen siete depósitos seleccionados aleatoriamente de cada sucursal, aproximado al US$100 más cercano. Hay $c = 4$ tratamientos (muestras) y $r_j = 7$ observaciones en cada tratamiento. El número total de observaciones es $n = rc = 28$.

Depósito	Sucursal 1	Sucursal 2	Sucursal 3	Sucursal 4
1	5.1	1.9	3.6	1.3
2	4.9	1.9	4.2	1.5
3	5.6	2.1	4.5	0.9
4	4.8	2.4	4.8	1.0
5	3.8	2.1	3.9	1.9
6	5.1	3.1	4.1	1.5
7	4.8	2.5	5.1	2.1
$\bar{X}_j$	4.87	2.29	4.31	1.46

$$\bar{\bar{X}} = \frac{\Sigma X_{ij}}{n} = \frac{(5.1 + 4.9 + 5.6 + \cdots + 2.1)}{28} = 3.23$$

Shade desea probar la hipótesis al nivel del 5% que

$$H_0\text{: } \mu_1 = \mu_2 = \mu_3 = \mu_4$$
$$H_A\text{: No todas las medias son iguales}$$

Utilizando las fórmulas (10.3) a (10.5) se tendría que

$$\begin{aligned} SCT &= \Sigma\Sigma(X_{ij} - \bar{\bar{X}})^2 \\ &= (5.1 - 3.23)^2 + (4.9 - 3.23)^2 + (5.6 - 3.23)^2 \\ &\quad + \cdots + (2.1 - 3.23)^2 \\ &= 61.00 \\ SCTR &= \Sigma\, r_j(\bar{X}_j - \bar{\bar{X}})^2 \\ &= 7(4.87 - 3.23)^2 + 7(2.29 - 3.23)^2 \\ &\quad + 7(4.31 - 3.23)^2 + 7(1.46 - 3.23)^2 \\ &= 55.33 \\ SCE &= \Sigma\Sigma(X_{ij} - \bar{X}_j)^2 \\ &= (5.1 - 4.87)^2 + \cdots + (4.8 - 4.87)^2 && \text{Para el primer tratamiento} \\ &\quad + (1.9 - 2.29)^2 + \cdots + (2.5 - 2.29)^2 && \text{Para el segundo tratamiento} \\ &\quad + (3.6 - 4.31)^2 + \cdots + (5.1 - 4.31)^2 && \text{Para el tercer tratamiento} \\ &\quad + (1.3 - 1.46)^2 + \cdots + (2.1 - 1.46)^2 && \text{Para el cuarto tratamiento} \\ &= 5.67 \end{aligned}$$

Las fórmulas (10.7) y (10.8) para los cuadrados medios da

$$CMTR = \frac{55.33}{3}$$
$$= 18.44$$
$$CME = \frac{5.67}{24}$$
$$= 0.236$$

Entonces la razón F es

$$F = \frac{CMTR}{CME}$$
$$= \frac{18.44}{0.236}$$
$$= 78.14$$

Shade debe utilizar 3 y 24 grados de libertad, ya que g.l. para $SCTR = 3$ y g.l. para $SCE = 24$. Si desea un α de 5%, encuentra en la tabla G (apéndice III) que $F_{0.05,3,24} = 3.01$. La tabla ANOVA resume estas cifras así

Fuente de variación	Suma de cuadrados	Grados de libertad	Cuadrados medios	Valor F
Entre muestras (tratamiento)	55.33	3	18.44	78.14
Dentro de muestras (error)	5.67	24	0.236	
Variación total	61.00	27		

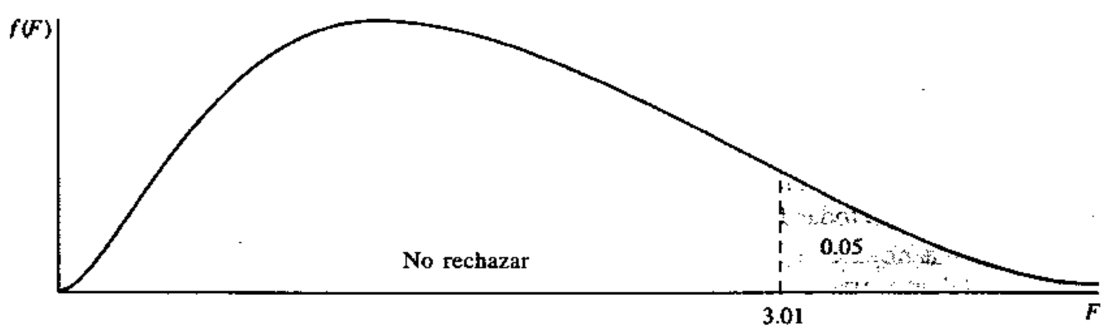
H_0: $\mu1 = \mu2 = \mu3 = \mu4$

H_A: No todas las medias son iguales

Regla de decisión: No rechazar si $F \leq 3.01$. Rechazar si $F > 3.01$

Conclusión: Debido a que $F = 78.14 > 3.01$, se rechaza la hipótesis nula.

La prueba se representa en la siguiente figura:



Interpretación

Debido a que $F = 78.14$, Shade debe rechazar la hipótesis nula. Puede estar 95% seguro de que los depósitos promedio en todas las sucursales bancarias no son iguales. Si considera que los grupos de ingreso diferentes se sienten atraídos por tipos de juegos de promoción distintos, debería diseñar esquemas alternativos para que cada sucursal atraiga nuevos depositantes.

10.3 Pruebas para la diferencia entre pares de medias

Como se puede observar en la explicación anterior, el análisis de varianza dice si todas las medias son iguales. Sin embargo, cuando se rechaza la hipótesis nula, el análisis de varianza no revela cuál(es) media(s) es (son) diferentes del resto. Se deben utilizar otras pruebas estadísticas para tomar esta determinación. Estas pruebas consisten en una comparación por pares, de todos los pares de medias posibles. Si el valor absoluto (ignorando los signos) de la diferencia entre dos medias muestrales cualquiera es mayor que algún estándar, se observa como una diferencia significativa, y se concluye que las medias poblacionales respectivas son diferentes.

Se puede determinar este estándar debido a una diversidad de procedimientos estadísticos incluyendo el método de Tukey (Too'Key) y la diferencia mínima significativa (*DMS*).

A. Pruebas para diseños balanceados

Tanto el método Tukey como el primero de los dos métodos *DMS* que aparecen aquí, se utilizan si existe igual número de observaciones en cada muestra. Se dice que tales diseños del análisis de varianza son *balanceados*. Si el diseño no está balanceado porque las muestras son de diferentes tamaños, debe utilizarse un método *DMS* alternativo (que se ilustrará en breve).

Diseños ANOVA En un diseño de análisis de varianza balanceado, cada muestra tiene el mismo número de observaciones. Si una o más muestras tienen un número diferente de observaciones, se dice que el diseño no está balanceado.

En el ejemplo 10.1, el Sr. Shade descubrió que no todas las cuatro sucursales de su banco tenían los mismos niveles de depósitos. El siguiente paso lógico es determinar cuáles son diferentes. Debido a que hay un número igual de observaciones en todas las cuatro muestras ($r = 7$), cualquiera de los métodos, bien sea el de Tukey o el DMS puede utilizarse.

El método Tukey Desarrollado en1953 por J.W. Tukey, requiere el cálculo del criterio de Tukey, T, como aparece en la fórmula (10.10).

Criterio de Tukey para comparaciones por pares $$T = q_{\alpha,c,n-c}\sqrt{\frac{CME}{r}} \qquad [10.10]$$

en donde q tiene una **distribución de rangos estudentizada** con c y $n - c$ grados de libertad y α es el valor α seleccionado. Vale la pena recordar que c es el número de muestras o tratamientos (columnas), y n es el número total de observaciones en todas las muestras combinadas. Estos valores son 4 y 28 en el problema de la sucursal bancaria de Shade.

La tabla L (apéndice III) proporciona los valores críticos para q con $\alpha = 0.01$ y $\alpha = 0.05$. Si α se fija en 0.05, Shade desea el valor para $q_{0.05,4,24}$. En la sección de la tabla L destinada a los valores con $\alpha = 0.05$, se pasa a la fila superior para los primeros grados de libertad de 4 y se baja por esa columna hasta los segundos grados de libertad de 24. Allí se encontrará el valor 3.90. Entonces

$$T = 3.90\sqrt{\frac{0.236}{7}}$$
$$= 0.716$$

El criterio estándar de Tukey de 0.716 se compara entonces con la diferencia absoluta entre cada par de medias muestrales. Si cualquier par de medias muestrales tiene una diferencia absoluta mayor que el valor T de 0.716, se puede concluir, a un nivel del 5%, que sus medias poblacionales respectivas no son iguales. La diferencia entre las medias muestrales es demasiado grande como para concluir que proviene de poblaciones similares. Existe sólo un 5% de probabilidad que las poblaciones con medias iguales puedan producir muestras de estos tamaños con medias que difieran en más de 0.716.

$$
\begin{aligned}
|\bar{X}_1 - \bar{X}_2| &= |4.87 - 2.29| = 2.58 > 0.716^* \\
|\bar{X}_1 - \bar{X}_3| &= |4.87 - 4.31| = 0.56 < 0.716 \\
|\bar{X}_1 - \bar{X}_4| &= |4.87 - 1.46| = 3.41 > 0.716^* \\
|\bar{X}_2 - \bar{X}_3| &= |2.29 - 4.31| = 2.02 > 0.716^* \\
|\bar{X}_2 - \bar{X}_4| &= |2.29 - 1.46| = 0.83 > 0.716^* \\
|\bar{X}_3 - \bar{X}_4| &= |4.31 - 1.46| = 2.85 > 0.716^*
\end{aligned}
$$

Al comparar los valores absolutos de cada diferencia entre los pares de medias muestrales con $T = 0.716$, Shade puede estar 95% seguro que sólo las sucursales 1 y 3 tienen igual nivel promedio de depósitos. Todas las otras diferencias exceden el criterio T.

Estos resultados pueden resumirse mediante el **subrayado común** en el cual las líneas que conectan las medias muestran que éstas *no* difieren significativamente. Las medias muestrales primero deben ponerse en un *serie ordenada*, generalmente del más bajo al más alto, tal y como se muestra aquí. Debido a que sólo las sucursales 1 y 3 no difieren significativamente, son las únicas que están conectadas por un subrayado común.

$$
\begin{array}{cccc}
\bar{X}_4 & \bar{X}_2 & \bar{X}_3 & \bar{X}_1 \\
\underline{1.46} & \underline{2.29} & \multicolumn{2}{c}{\underline{4.31 \quad\quad 4.87}}
\end{array}
$$

Diferencia mínima significativa El método de diferencia mínima significativa es muy similar al método de Tukey. Compara el criterio de la diferencia menos significativa con la diferencia absoluta en las medias muestrales.

Si el diseño está balanceado, el criterio DMS es:

Diferencia mínima significativa
$$DMS = \sqrt{\frac{2(CME)F_{\alpha,1,n-c}}{r}} \qquad [10.11]$$

Vale la pena destacar que al utilizar el método DMS, F tiene 1 y $n - c$ grados de libertad. En el caso de Shade esto es 1 y $n - c = 28 - 4 = 24$ grados de libertad. De la tabla F, $F_{0.05,1,24} = 4.26$. Entonces:

$$
\begin{aligned}
DMS &= \sqrt{\frac{2(0.236)4.26}{7}} \\
&= 0.536
\end{aligned}
$$

Al comparar la *DMS* de 0.536 con cada una de las diferencias absolutas que aparecieron anteriormente, Shade encuentra que todos los valores, incluyendo el último, sugiere medias poblacionales diferentes. El método *DMS* es más conservador en que, dado un conjunto de condiciones cualquiera, el criterio *DMS* será menor que el valor Tukey.

Los cálculos matemáticos extensivos que necesitó el análisis de varianza puede facilitarse con el uso de paquetes de software modernos. La pantalla 10.1 muestra la impresión del ejemplo 10.1 en el cual el Sr. Shade del First City Bank tenía que decidir si los depósitos promedio en cuatro sucursales bancarias eran iguales. La porción superior muestra el valor F de 78.09, el cual se compara con el 78.14 que se calculó manualmente. El valor p de 0.000 revela por qué se rechazó la hipótesis nula a un nivel del 5%.

La parte inferior de la impresión proporciona el subrayado común. De acuerdo con el criterio de Tukey, se puede observar que sólo las sucursales 1 y 3 se superponen.

Pantalla 10.1

One-Way Analysis of Variance (Análisis de varianza a una sola vía)

```
Analysis of Variance on C1 (Análisis de varianza en C1)  DF     = Grados de libertad
Source      DF        SS       MS       F        P        SS     = Sumas de cuadrados
C2           3    55.333   18.444   78.09    0.000        MS     = Cuadrado medio
Error       24     5.669    0.236                         Source = Fuente
Total       27    61.001
                                           Individual 95% CIs For Mean Based on Pooled
                                           StDev (ICs del 95% para media basada en desvia-
Level    N    Mean                 StDev   ción estándar mancomunada)
(Nivel)     (Media) (Desviación estándar)  --+---------+---------+--------+----
1        7  4.8714                0.5469                                (---*--)
2        7  2.2857                0.4259          (--*--)
3        7  4.3143                0.5210                           (--*--)
4        7  1.4571                0.4392     (--*--)
                                           --+---------+---------+---------+----
Pooled StDev (Desviación
estándar mancomunada) = 0.4860              1.2       2.4       3.6       4.8
MTB >
```

Con cualquiera de los métodos, pueden surgir inconsistencias. Se asume en aras de la simplicidad que hay sólo tres poblaciones en el estudio, que requieren comparaciones por pares:

$$|\bar{X}_1 - \bar{X}_2| \qquad |\bar{X}_1 - \bar{X}_3| \qquad |\bar{X}_2 - \bar{X}_3|$$

Puede encontrarse que 1 no difiere significativamente de 2, y que 2 no difiere significativamente de 3, pero que 1 sí difiere significativamente de 3. Esto parece contradictorio. Pero según la regla de la transitividad, si 1 es igual a 2 y 2 es igual a 3, entonces 1 debe ser igual a 3. Sin embargo, las comparaciones por pares no involucran igualdades. Al comparar las tres poblaciones, se analiza simplemente la evidencia estadística para determinar si es lo suficientemente fuerte como para rechazar la hipótesis nula. Concluir que 1 no difiere significativamente de 2 simplemente significa que se tiene suficiente evidencia para concluir que son diferentes. Si se concluye, como se hizo aquí, que 1 difiere de 3, puede asumirse que la evidencia que compara estas dos muestras era más fuerte.

B. Pruebas para diseños no balanceados

Si el diseño es no balancedo, el método de Tukey y el método *DMS*, discutidos anteriormente, simplemente no se aplican, en su lugar, se puede utilizar un método *DMS* alternativo.

Método *DMS* alternativo Para comparar las muestras *jésima* y *késima*, la ecuación para DMS se vuelve:

Diferencia mínima significativa para el diseño no balanceado

$$DMS_{j,k} = \sqrt{\left[\frac{1}{r_j} + \frac{1}{r_k}\right](CME)F_{\alpha,c-1,n-c}} \qquad [10.12]$$

en donde r_j es el número de observaciones en la muestra *jésima* y r_k es el número de observaciones en la muestra *késima*. El valor *DMS* será diferente para cada par de comparaciones por pares, debido a que el número de observaciones no es el mismo en cada muestra.

Ejemplo 10.2

Cada vez más norteamericanos buscan escapar de las presiones urbanas, los pagos impuestos en los parques nacionales ha demostrado un incremento marcado de quienes acampan los fines de semana. *Outdoor World* informó recientemente que el parque Yosemite National Park ubicado en las sierras altas de California contrató un consultor en economía para estudiar la situación financiera del parque.

Parte del esfuerzo realizado por el consultor requería una comparación de los ingresos del parque provenientes de varias fuentes, incluyendo los pagos por acampar, licencias para pescar y para pasear en bote. Aquí aparecen los datos para visitantes, seleccionados aleatoriamente. Se determina si existe diferencia en los ingresos promedio que recibe el parque provenientes de estas tres actividades.

Visitante	Acampar	Pesca	Pasear en bote
1	US$38.00	US$30.00	US$19.00
2	32.00	25.00	35.00
3	35.00	31.00	20.00
4	36.00	35.00	22.00
5	38.00	**	25.00
6	32.00	**	**
$\bar{X}_j$	US$35.17	US$30.25	US$24.20

Solución

Asumiendo que α se fija en 5%, entonces $F_{\alpha,c-1,n-c} = F_{0.05,2,12} = 3.89$. La tabla ANOVA aparecería así:

Fuente de variación	Suma de cuadrados	Grados de libertad	Cuadrado medio	Valor F
Entre muestras (tratamiento)	328.0	2	164.0	7.74
Dentro de muestras (error)	254.4	12	21.2	
Variación total	582.4	14		

H_0: $\mu 1 = \mu 2 = \mu 3$

H_A: No todas las medias son iguales

Regla de dicisión: No rechazar si $F \leq 3.89$. Rechazar si $F > 3.89$

Conclusión: Rechazar la hipótesis nula ya que $F = 7.74 > 3.89$.

Debido a que se rechaza la hipótesis nula de los ingresos promedio provenientes de todas las tres actividades, el consultor desearía utilizar las comparaciones por pares para determinar cuáles difieren del resto. Si α es 5%, $F_{0.05,c-1,n-c} = F_{0.05,2,12} = 3.89$. La comparación para la primera actividad (acampar) y para la segunda (pesca), utilizando la fórmula (10.2) para calcular *DMS* es:

$$DMS_{C,F} = \sqrt{\left[\frac{1}{6} + \frac{1}{4}\right](21.2)(3.89)}$$
$$= 5.85$$

Una comparación entre acampar y pasear en bote revela que:

$$DMS_{C,B} = \sqrt{\left[\frac{1}{6} + \frac{1}{5}\right](21.2)(3.89)}$$
$$= 5.48$$

La última comparación entre pesca y montar en bote produce

$$DMS_{F,B} = \sqrt{\left[\frac{1}{4} + \frac{1}{5}\right](21.2)(3.89)}$$
$$= 6.08$$

Las diferencias entre las medias y si exceden o no su valor *DMS* respectivo son

$$
\begin{aligned}
|\overline{X}_c - \overline{X}_f| &= |35.17 - 30.25| = 4.92 < 5.85 \\
|\overline{X}_c - \overline{X}_b| &= |35.17 - 24.20| = 10.97 > 5.48 \\
|\overline{X}_f - \overline{X}_b| &= |30.25 - 24.20| = 6.05 < 6.08
\end{aligned}
$$

Sólo acampar y montar en bote difieren significativamente. Los resultados pueden resumirse con un subrayado común después que las medias se hayan colocado en un arreglo ordenado así.

$$
\begin{array}{ccc}
\overline{X}_b & \overline{X}_f & \overline{X}_c \\
24.2 & \underline{30.25 \quad 35.17} & \\
\end{array}
$$

Interpretación

Se puede concluir a un nivel de significancia del 5% que sólo montar en bote y acampar difieren significativamente. El parque puede utilizar esta información para tomar decisiones y aliviar la presión financiera sobre los recursos y proporcionar una experiencia al aire libre para los pioneros modernos.

Parece que el subrayado común del ejemplo es auto-contradictorio. Muestra que montar en bote y pescar no difieren y que la pesca y el hecho de acampar no son diferentes, sin embargo montar en bote y acampar sí son diferentes. La regla algebraica de la transitividad dice que si *A* igual a *B* y *B* igual a *C*, ¿entonces *A* debe ser igual a *C*? Sí, pero no se está tratando de igualdades con este ejemplo. Simplemente se dice que la diferencia entre montar en bote y pescar no es significativa y que la diferencia entre pescar y acampar no es significativa, sin embargo la diferencia entre montar en bote y acampar es lo suficientemente grande como para ser significativa.

10.4 ANOVA a dos vías. El diseño aleatorizado en bloques

Con el análisis de varianza a una vía, se pensó que sólo un factor influenciaba las unidades experimentales – tal como los depósitos en las sucursales bancarias, o los ingresos en el parque. Sin embargo, con frecuencia se encuentra que una segunda influencia *exterior* puede impactar las unidades experimentales. Por ejemplo, el interés puede ser comparar la productividad promedio de los tres tipos de máquinas (tratamientos). Sin embargo, se observa que al probar estas máquinas, la destreza del operador y su experiencia pueden afectar la producción de la máquina, produciendo confusión sobre cuál máquina es realmente mejor. Así, para obtener un panorama no contaminado y claro de la capacidad de la máquina, se debe eliminar de alguna manera o corregir, la influencia del operador sobre la producción final. Esta consideración simultánea de las dos fuerzas requiere del **análisis de varianza a dos vías.**

Para obtener una medida decisiva de la capacidad de la máquina, se debe "bloquear" el factor externo, colocando las observaciones en grupos homogéneos con base en los años de experiencia. Así, las observaciones se clasifican tanto por bloques como por tratamientos. El propósito del bloqueo es reducir la variación dentro de un tratamiento (tipo de máquina). Este diseño experimental se llama **diseño aleatorizado en bloques**.

Si los bloques se realizan de manera efectiva y se basan en un factor (tal como la experiencia) que verdaderamente afecte la productividad, se obtiene una medida más pura del efecto del tratamiento. Sin embargo, si el factor seleccionado para el bloqueo no afecta la productividad (como por ejemplo el número de seguro social del trabajador, el color del cabello o el sexo), los resultados pueden ser engañosos. Es importante determinar si el bloqueo se hace o no correctamente, y si el factor en el que se basa el bloqueo sí tiene cierto impacto.

Para ilustrar, una empresa de contabilidad grande trata de seleccionar un sistema de computación integrado a la oficina, entre los tres modelos que están actualmente en estudio. La selección final dependerá de la productividad de los sistemas. Se seleccionan aleatoriamente cinco operadores para manejar cada sistema. Es importante tener en cuenta que el nivel de experiencia que tienen los empleados en el manejo de computadores puede afectar el resultado de la prueba. Por tanto, existe la necesidad de justificar el impacto de la experiencia al determinar los méritos relativos de los sistemas de computación. Los niveles resultantes de producción medidos en unidades por hora aparecen en la tabla 10.3. Un valor codificado más alto para la experiencia indica más años de capacitación.

Tabla 10.3
Niveles de producción para los sistemas de computación

	Sistemas (tratamientos)			
Nivel de experiencia	**1**	**2**	**3**	$\bar{X}_i$
1	27	21	25	24.33
2	31	33	35	33.00
3	42	39	39	40.00
4	38	41	37	38.67
5	45	46	45	45.33
$\bar{X}_i$	36.5	36.0	36.2	
				$\bar{\bar{X}} = 36.27$

Dentro de una muestra dada (sistema) ocurrirá una variación en la producción debido a la experiencia del operador, su competencia y su estado actual de salud, y a otros factores de error aleatorios. En el análisis de varianza a una vía, se identificó como error de variación. Si cualquiera de estos factores aleatorios relacionados con los operadores afectan materialmente el nivel de producción, la empresa de contabilidad debe corregirlos. La empresa puede considerar que los años de experiencia de un operador afectan significativamente su productividad. Sin embargo, la empresa está interesada en la productividad de los sistemas de computación, y no en la de los empleados. Por tanto se debe ajustar a la productividad de los empleados eliminando el efecto de variabilidad del operador para obtener una medida precisa, no contaminada, de la calidad del sistema.

Con el análisis de varianza a dos vías, la suma de cuadrados total se divide en tres partes: la suma de cuadrados del tratamiento (*SCTR*), suma de cuadrados del error, y la suma de cuadrados de bloques (*SCBL*). Por tanto

$$SCT = SCTR + SCE + SCBL$$

SCT y *SCTR* se calculan de la misma forma que en el análisis de varianza a una vía. Sin embargo *SCE* se subdivide en una medida para *SCE* y *SCBL*, en donde

Suma de cuadrados de bloques	$SCBL = \Sigma c_i(\bar{X}_i - \bar{\bar{X}})^2$	[10.13]

El número de tratamientos en cada bloque , c_i, se multiplica por la diferencia al cuadrado entre la media para cada bloque, $\bar{X}_i$, y la gran media. Los resultados se suman para todos los bloques. El símbolo c_i se utiliza para indicar el número de tratamientos en un bloque (fila), porque los tratamientos se registran en las columnas. De la tabla 10.3,

$$\begin{aligned} SCBL &= 3(24.33 - 36.27)^2 + 3(33 - 36.27)^2 + 3(40 - 36.27)^2 \\ &\quad + 3(38.67 - 36.27)^2 + 3(45.33 - 36.27)^2 \\ &= 765.04 \end{aligned}$$

La suma de los cuadrados del bloque mide el grado de variación de las medias del bloque (fila) alrededor de la gran media.

Las fórmulas (10.3) y (10.4) dan

$$SCT = 806.93 \qquad \text{y} \qquad SCTR = 0.93$$

SCE se calcula como

Suma de cuadrados del error	$SCE = SCT - SCTR - SCBL$	[10.14]

$$\begin{aligned} &= 806.93 - 0.93 - 765.04 \\ &= 40.96 \end{aligned}$$

En donde hay r bloques y c tratamientos, y hay $n = rc$ observaciones. Los grados de libertad para cada una de las sumas de cuadrados de los valores de la fórmula (10.14) son

$$\begin{array}{ccccccc} SCE & = & SCT & - & SCTR & - & SCBL \\ (r-1)(c-1) & = & (n-1) & - & (c-1) & - & (r-1) \\ (5-1)(3-1) & = & (15-1) & - & (3-1) & - & (5-1) \\ 8 & = & 14 & - & 2 & - & 4 \end{array}$$

El cuadrado medio total y el cuadrado medio del tratamiento son, al igual que antes, la suma de sus cuadrados dividida por sus grados de libertad. Por tanto,

$$\text{Cuadrado medio total} = CMT = \frac{SCT}{n-1}$$

$$= \frac{806.93}{14}$$

$$= 57.64$$

$$\text{Cuadrado medio del tratamiento} = CMTR = \frac{SCTR}{c-1}$$

$$= \frac{0.93}{2}$$

$$= 0.47$$

En ANOVA a dos vías,

$$\text{Cuadrado medio del error} \qquad CME = \frac{SCE}{(r-1)(c-1)} \qquad [10.15]$$

$$= \frac{40.96}{8}$$

$$= 5.1$$

$$\text{Cuadrado medio del bloque} \qquad CMBL = \frac{SCBL}{r-1} \qquad [10.16]$$

$$= \frac{765.04}{4}$$

$$= 191.26$$

Tabla 10.4
ANOVA a dos vías para los sistemas de computación

Fuente de variación	Suma de cuadrados	Grados de libertad	Cuadrado medio	Valor F
Entre muestras (tratamiento)	0.93	2	0.47	0.09
Entre bloques	765.04	4	191.26	37.50
Dentro de las muestras (error)	40.96	8	5.10	
Variación total	806.93	14		

Estos cálculos se resumen en la tabla 10.4. Los valores F se calculan de la misma manera que en el análisis de varianza a una vía:

$$F = \frac{CMTR}{CME}$$

$$= \frac{0.47}{5.1}$$

$$= 0.09$$

$$F = \frac{CMBL}{CME}$$

$$= \frac{191.26}{5.1}$$

$$= 37.50$$

Vale la pena notar que se calculan dos valores F – uno que utiliza *CMTR* y uno que utiliza *CMBL*. El valor F para *CMBL* se calcula para determinar si los bloques se realizaron de manera efectiva. Si el bloqueo se basa en un factor que no afecta la productividad del operador, los resultados pueden ser engañosos. Por tanto, la empresa de contabilidad debe hacer pruebas para ver si existe una diferencia significativa entre las medias del bloque (fila). Si no existe diferencia significativa entre los niveles promedio de producción con base en los bloques (filas), entonces la experiencia no es un factor crítico. En este caso, se debería abandonar el análisis de varianza a dos vías, y se necesitaría regresar al análisis de varianza a una sola vía, sin distinción entre los niveles de experiencia. A un nivel del 5%, el valor crítico de F para *CMBL* con 4 y 8 grados de libertad se obtiene de la tabla G y es $F_{0.05,4,8} = 3.84$. Los grados de libertad 4 y 8 se utilizan porque la relación F para los bloques utiliza *CMBL* con $r - 1 = 4$ grados de libertad y *CME* con $(r - 1)(c - 1) = 8$ grados de libertad.

La firma contable debe primero probar la hipótesis de que el nivel promedio de producción para cada nivel de experiencia es el mismo. Si es así, entonces la experiencia no es un factor determinante en la producción, y el bloqueo sobre ésta sería inútil. Si los niveles promedio de producción y los niveles de experiencia no son los mismos, entonces la empresa de contabilidad debe bloquear la experiencia para corregir su impacto y por ende obtener una medida más exacta de las diferencias en la calidad del sistema de computación. La hipótesis a probar es

$$H_0: \mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5$$

H_A: No todas las medias de las filas son iguales

en donde μ_i son los niveles promedio de producción para cada nivel de experiencia (fila).

Regla de decisión: "No rechazar la hipótesis nula si $F \leq 3.84$. Rechazar la hipótesis nula si $F > 3.84$".

Debido a que $F = 37.50$, la hipótesis nula debería rechazarse, y la empresa debería concluir que los niveles de experiencia tienen un efecto en las tasas de producción. Debe corregir la experiencia utilizando el análisis de varianza a dos vías.

Ahora la empresa está preparada para probar la hipótesis en la cual estuvo originalmente interesada. ¿Existe alguna diferencia en la producción promedio de los sistemas de computación (tratamientos)? Si el valor α del 5% se mantiene, $F_{\alpha,(c-1),(r-1)(c-1)} = F_{0.05,2,8} = 4.46$ se obtiene de la tabla. Los grados de libertad de 2 y 8 se utilizan porque la razón F para los tratamientos utiliza *CMTR* con 2 grados de libertad y *CME* con 8 grados de libertad. El conjunto de hipótesis es

$$H_0: \mu_1 = \mu_2 = \mu_3$$

H_A: No todas las medias de las columnas son iguales

En donde μ_i son las medias de las columnas para los tres sistemas de computación.

Regla de decisión: "No rechazar la hipótesis nula si $F \leq 4.46$. Rechazar la hipótesis nula si $F > 4.46$".

La tabla 10.4 indica que $F = 0.09 < 4.46$. La hipótesis nula no se rechaza, y la empresa concluye que los niveles de producción promedio de los tres sistemas de computación no difieren, una vez que se ha hecho la corrección para el factor experiencia. Los empleados de diferentes niveles de experiencia se desempeñan igualmente bien en todas las máquinas. No interesa cuál sistema de computación compren.

El ejemplo 10.3 proporciona otra ilustración del ANOVA o análisis de varianza a dos vías

Ejemplo 10.3

Una emisión reciente de la revista *Fortune* describió los esfuerzos realizados por una importante empresa de electrónica para desarrollar un sistema en el cual se les daba a los empleados la oportunidad de evaluar el desempeño de sus supervisores y de algún personal administrativo. Se seleccionan aleatoriamente cinco empleados y se les pide evaluar a cuatro de sus gerentes sobre una escala de 10 a 50. Los resultados, junto con las medias de las filas y las columnas, aparecen en la siguiente tabla.

	Gerente (tratamiento)				
Empleado	**1**	**2**	**3**	**4**	$\bar{X}_i$
1	31	35	46	38	37.50
2	29	32	45	36	35.50
3	13	17	35	20	21.25
4	28	38	52	39	39.25
5	14	20	40	20	23.50
$\bar{X}_j$	23	28.4	43.6	30.6	$\bar{\bar{X}} = 31.4$

El gerente de la empresa de electrónica desea saber si existe diferencia en las clasificaciones promedio de los cuatro gerentes.

Solución

El director decide utilizar el análisis de varianza a dos vías para probar las medias:

$$
\begin{aligned}
SCT &= \Sigma\Sigma(X_{ij} - \bar{\bar{X}})^2 \\
&= (31 - 31.4)^2 + (29 - 31.4)^2 + \cdots + (39 - 31.4)^2 \\
&\quad + (20 - 31.4)^2 \\
&= 2344.8 \\
SCTR &= \Sigma r_j(\bar{X}_j - \bar{\bar{X}})^2 \\
&= 5(23 - 31.4)^2 + 5(28.4 - 31.4)^2 + 5(43.6 - 31.4)^2 \\
&\quad + 5(30.6 - 31.4)^2 \\
&= 1145.2 \\
SCBL &= \Sigma c_i(\bar{X}_i - \bar{\bar{X}})^2 \\
&= 4(37.5 - 31.4)^2 + 4(35.5 - 31.4)^2 \\
&\quad + 4(21.25 - 31.4)^2 + 4(39 - 31.4)^2 \\
&\quad + 4(23.5 - 31.4)^2 \\
&= 1124.3 \\
SCE &= SCT - SCTR - SCBL \\
&= 2344.8 - 1145.2 - 1124.3 \\
&= 75.3
\end{aligned}
$$

La tabla ANOVA a dos vías se convierte en:

Fuente de variación	Suma de cuadrados	Grados de libertad	Cuadrado medio	Valor F
Entre muestras (tratamiento)	1,145.2	3	381.73	60.79
Entre bloques	1,124.3	4	281.08	44.76
Dentro de muestras (error)	75.3	12	6.28	
Variación total	2,344.8	19		

Ahora el director puede determinar si hay una diferencia significativa en las clasificaciones promedio dadas por cada uno de los cinco empleados (filas), las cuales requerirán hacer bloques sobre los empleados. Las hipótesis son:

$$H_0\text{: } \mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5$$

H_A: No todas las medias de las filas son iguales

Si $\alpha = 1\%$, el valor F apropiado es $F_{0.01,4,12} = 5.41$. El valor F relacionado con la prueba sobre los bloques aparece en la tabla de análisis de varianza como 44.76> 5.41. La hipótesis nula se rechaza, y el director determina, a un nivel de significancia del 1%, que las clasificaciones promedio hechas por los cinco empleados (filas) son diferentes y se necesita el bloqueo.

El director puede probar ahora su hipótesis inicial respecto a las clasificaciones promedio de los cuatro gerentes (columnas). Las hipótesis son

$$H_0\text{: } \mu_1 = \mu_2 = \mu_3 = \mu_4$$

H_A: No todas las medias de las columnas son iguales

El valor F de $F_{0.01,3,12} = 5.95$ es menor que 60.79. La hipótesis nula debe rechazarse a un nivel de significancia del 1%.

Interpretación

Al incluir un factor de bloqueo, el director pudo detectar una diferencia significativa en la clasificación promedio de los gerentes realizada por parte de los cinco empleados. Sin el factor de bloqueo, la variación en las clasificaciones, debido a los bloques (diferencias en las actitudes de los empleados), se hubiera incluido en el factor de error *SCE*. Esto hubiera tenido el efecto de incrementar la *SCE* y el *CME*.

El valor *F*, por tanto, hubiese sido menor debido a que *F* = *CMTR/CME*. A medida que el valor *F* disminuye, existe una mayor probabilidad de no rechazar la hipótesis nula.

Sin embargo, con el análisis de varianza a dos vías, el *CME* se subdivide en la variación debida a los bloques (*CMBL*) y en la variación debida al error dentro de las muestras (*CME*).

Ahora que el director sabe que no todas las clasificaciones de los gerentes son las mismas, puede utilizar el método de Tukey o el método **DMS** para determinar cuáles son diferentes. Al aplicar estas herramientas a una prueba a dos vías, deben hacerse ciertos cambios en los grados de libertad relacionados con el método Tukey. En lugar de explorar ese ajuste, el método **DMS** puede utilizarse con el análisis de varianza a dos vías como se demostró anteriormente.

La pantalla 10.2 muestra la impresión en Minitab del ejemplo 10.3. La porción superior proporciona la tabla de análisis de varianza. Sin embargo, como lo expresa el manual de Minitab, no se puede especificar si los efectos son aleatorios o fijos con el comando **TWOWAY**. Por tanto, Minitab no proporciona los valores F o los valores p. Minitab aparentemente espera que usted los calcule manualmente. La hipótesis para las medias de las filas (bloqueo sobre los empleados)

$$H_0\text{: } \mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5$$

lleva un valor F de

$$\frac{CMBL}{CME} = \frac{281.08}{6.28} = 44.76 > F_{0.05,4,12} = 3.26$$

La hipótesis nula de medias iguales de la fila se rechaza al nivel del 5%, y se utiliza el análisis de varianza a dos vías para probar la hipótesis primaria de iguales clasificaciones promedio para los gerentes:

$$H_0\text{: } \mu_1 = \mu_2 = \mu_3 = \mu_4$$

El valor F es

$$\frac{CMTR}{CME} = \frac{381.73}{6.28} = 60.79 > F_{0.05,3,12} = 3.49$$

La hipótesis nula de clasificaciones promedio para los gerentes se rechaza al nivel del 5%. Las comparaciones por pares se muestran en las porciones restantes de la impresión.

Los resultados de la prueba para las medias de la fila (empleado) se hallan en la sección del medio de la impresión. Vale la pena notar que no todas las medias son iguales. Las medias de los empleados 3 y 5 son menores que las otras tres. Los resultados para la hipótesis primaria respecto a las clasificaciones de los gerentes se encuentran en la última porción de la impresión. El gerente 3 tiene la clasificación más alta y el gerente 1 tiene la clasificación más baja. Al nivel de significancia del 5% no parece haber ninguna diferencia en las clasificaciones promedio de los gerentes 2 y 4.

Pantalla 10.2

Two-Way Analysis of Varience (Análisis de varianza a dos vías)

```
Analysis of Varience for rating
Source (Fuente(          DF         SS          MS          DF = Grados de libertad
Employee (Empleado)       4    1124.30      281.08          SS = Sumas de cuadrados
Manager (Gerente)         3    1145.20      381.73          MS = Cuadrado medio
Error                    12      75.30        6.28
Total                    19    2344.80
                                      Individual 95% CI (I.C. del 95% individual)
Employee (Empleado) Mean (Media)      ----------+---------+----------+---------+-
              1            37.5                                    (----*---)
              2            35.5                                 (---*----)
              3            21.2         (---*----)
              4            39.2                                         (---*----)
              5            23.5              (---*----)
                                      ----------+---------+----------+---------+-
                                              24.0      30.0       36.0      42.0
                                      Individual 95% CI (I.C. del 95% individual)
Manager (Gerente)    Mean (Media)     -+---------+---------+----------+---------+
              1            23.0        (---*--)
              2            28.4                  (---*--)
              3            43.6                                     (---*--)
              4            30.6                      (---*--)
                                       -+--------+---------+----------+---------+-
                                     21.0      28.0      35.0       42.0      49.0
MTB >
```

Lista de fórmulas

[10.1]	$\bar{\bar{X}} = \frac{\Sigma X_{ij}}{n}$	Gran media de todas las observaciones del experimento
[10.2]	$s^2 = \frac{\Sigma(X_i - \bar{X})^2}{n-1}$	La varianza muestral
[10.3]	$SCT = \sum_{i=1}^{r}\sum_{j=1}^{c}(X_{ij} - \bar{\bar{X}})^2$	La suma de cuadrados total es la variación de los valores alrededor de $\bar{\bar{X}}$
[10.4]	$SCTR = \Sigma r_j(\bar{X}_j - \bar{\bar{X}})^2$	La suma de los cuadrados del tratamiento mide la variación en las medias de la columna alrededor de la gran media
[10.5]	$SCE = \Sigma\Sigma(X_{ij} - \bar{X}_j)^2$	La suma de los cuadrados del error refleja la variación de los valores dentro de un tratamiento alrededor de su propia media
[10.6]	$CMT = \frac{SCT}{n-1}$	Cuadrado medio total
[10.7]	$CMTR = \frac{SCTR}{c-1}$	Cuadrado medio del tratamiento
[10.8]	$CME = \frac{SCE}{n-c}$	Cuadrado medio del error
[10.9]	$F = \frac{CMTR}{CME}$	Razón F para la prueba de las medias
[10.10]	$T = q_{\alpha,c,n-c}\sqrt{\frac{CME}{r}}$	El criterio de Tukey que mide el valor crítico de la diferencia entre medias
[10.11]	$DMS = \sqrt{\frac{2(CME)F_{\alpha,1,n-c}}{r}}$	El criterio *DMS* mide la diferencia crítica entre dos medias - para diseños balanceados
[10.12]	$DMS_{j,k} = \sqrt{\left[\frac{1}{r_j} + \frac{1}{r_k}\right](CME)F_{\alpha,1,n-c}}$	El criterio *DMS* mide la diferencia crítica entre dos medias - para diseños no balanceados.
[10.13]	$SCBL = \Sigma c_i(\bar{X}_i - \bar{\bar{X}})^2$	La suma de los cuadrados del bloque mide las desviaciones de las medias de la fila alrededor de la gran media
[10.14]	$SCE = SCT - SCTR - SCBL$	La suma de cuadrados del error mide la variación aleatoria de las observaciones alrededor de sus medias de tratamiento
[10.15]	$CME = \frac{SCE}{(r-1)(c-1)}$	Cuadrado medio del error en análisis de varianza a dos vías
[10.16]	$CMBL = \frac{SCBL}{r-1}$	Cuadrado medio de bloque

Problemas resueltos

1. **Fleecing the Motorist**, *Estudio sobre los consumidores,* publicó los resultados de una encuesta sobre los hábitos de conducción en los Estados Unidos. Los datos contenían impuestos a la gasolina por hogar para los 50 estados. Seis estados se escogieron aleatoriamente de cuatro regiones del país para determinar si hay alguna diferencia en el impuesto promedio anual de la gasolina dentro de las regiones. Los siguientes resultados se aproximaron al dólar más próximo:

	Región (tratamiento)			
Estado	**Norte(1)**	**Sur(2)**	**Occidente(3)**	**Centro de EEUU(4)**
1	US$293	US$121	US$114	US$136
2	280	116	176	164
3	283	223	224	117
4	242	238	183	153
5	268	118	159	152
6	184	222	149	108
$\overline{X}_j$	258.3	173.0	167.5	138.3
$\overline{\overline{X}} = 184.3$				

Un economista deseaba probar al nivel del 5% la hipótesis de que en promedio los residentes de las cuatro regiones pagan la misma suma en los impuestos federales sobre la gasolina.

Solución

$$
\begin{aligned}
SCT &= \Sigma\Sigma(X_{ij} - \overline{\overline{X}})^2 \\
&= (293 - 184.3)^2 + \cdots + (108 - 184.3)^2 \\
&= 83{,}515
\end{aligned}
$$

SCTR se halla:

$$
\begin{aligned}
SCTR &= \Sigma r_j(\overline{X}_j - \overline{\overline{X}})^2 \\
&= 6(258.3 - 184.3)^2 + \cdots + 6(138.3 - 184.3)^2 \\
&= 48{,}023
\end{aligned}
$$

SCE es:

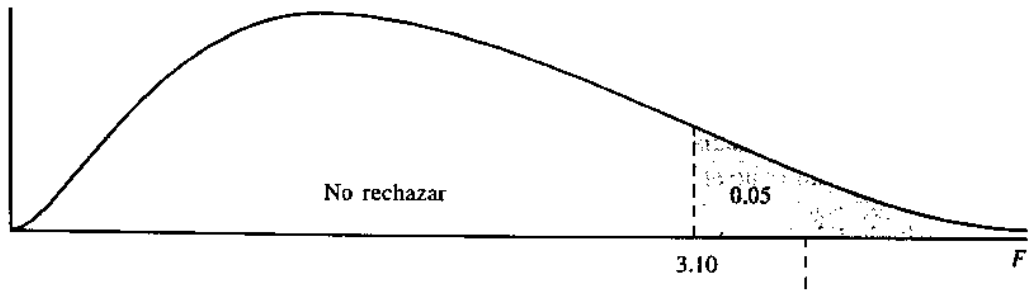
$$
\begin{aligned}
SCE &= \Sigma\Sigma(X_{ij} - \overline{X}_j)^2 \\
&= (293 - 258.3)^2 + \cdots + (108 - 138.3)^2 \\
&= 35{,}492
\end{aligned}
$$

Entonces,

$$
\begin{aligned}
CMTR &= \frac{CSTR}{c - 1} \\
&= \frac{48{,}023}{4 - 1} \\
&= 16{,}008
\end{aligned}
$$

$$
\begin{aligned}
CME &= \frac{SCE}{n - c} \\
&= \frac{35{,}492}{24 - 4} \\
&= 1{,}775
\end{aligned}
$$

Si α se fija en 5%, $F_{0.05,3,20} = 3.10$ como se observa aquí. La hipótesis es H_0: $\mu_1 = \mu_2 = \mu_3 = \mu_4$.



La tabla del análisis de varianza es

Fuente de variación	SC	g.l.	CM	Valor F
Entre muestras (tratamiento)	48,023	3	16,008	9.02
Dentro de muestras (error)	35,492	20	1,775	
Variación total	83,515	23		

H_0: $\mu_1 = \mu_2 = \mu_3 = \mu_4$

H_A: No todas las medias son iguales

Regla de decisión: No rechazar la hipótesis nula si $F \leq 3.10$. Rechazar si $F > 3.10$

Conclusión: rechazar la hipótesis nula

Puede concluirse a un nivel de significancia del 5% que el impuesto promedio no es el mismo en las cuatro regiones.

2. **¿A quiénes se les cobra más duro**? Para formular un sistema tributario efectivo, el gobierno debe determinar cuáles regiones pagan más y cuáles pagan menos. Utilizando tanto el método de Tukey como el método *DMS*, los cálculos se realizarían como se muestra aquí.

Solución

Primero es necesario hallar las diferencias absolutas entre las medias muestrales de los impuestos pagados en cada par de las cuatro regiones.

$$
\begin{aligned}
|\bar{X}_1 - \bar{X}_2| &= |258.3 - 173.0| = 85.3 \\
|\bar{X}_1 - \bar{X}_3| &= |258.3 - 167.5| = 90.8 \\
|\bar{X}_1 - \bar{X}_4| &= |258.3 - 138.3| = 120.0 \\
|\bar{X}_2 - \bar{X}_3| &= |173.0 - 167.5| = 5.5 \\
|\bar{X}_2 - \bar{X}_4| &= |173.0 - 138.3| = 34.7 \\
|\bar{X}_3 - \bar{X}_4| &= |167.5 - 138.3| = 29.2
\end{aligned}
$$

El criterio de Tukey es:

$$T = q_{\alpha,c,n-c}\sqrt{\frac{CME}{r}}$$

Si α se fija en 5%, $q_{0.05,4,20} = 3.96$,

$$T = 3.96\sqrt{\frac{1,775}{6}}$$

$$= 68.11$$

Toda diferencia absoluta entre las medias muestrales mayor que 68.11 es significativa y sugiere que sus respectivas medias poblacionales son diferentes. Existe sólo un 5% de probabilidad de que dos de estas poblaciones puedan tener la misma media y generar muestras de estos tamaños con medias superiores a 68.11. Al comparar 68.11 con los seis pares de las medias muestrales anteriores, puede observarse que la población 1 (norte) tiene una media diferente de las otras tres; se asume que μ_1 es más alto debido a que $\overline{X}_1$ es significativamente más alto que el resto.

Utilizando el método *DMS* se tiene que:

$$DMS = \sqrt{\frac{2(CME)F_{\alpha,1,n-c}}{r}}$$

$F_{0.05,1,20} = 4.35$ Entonces,

$$DMS = \sqrt{\frac{2(1,775)(4.35)}{6}}$$

$$= 50.73$$

El criterio *DMS* se compara con las diferencias absolutas entre las medias muestrales anteriores. Toda diferencia mayor que 50.73 es significante. De nuevo se observa que los motoristas del norte tienen la carga tributaria más pesada.

3. **Acme Ltd.** Como supervisor de producción de Acme Ltd., Melvin Moore desea comparar los niveles de producción de las cuatro plantas de Acme. Los datos semanales sobre los niveles en toneladas se recolectan durante un período determinado de siete semanas, incluyendo las cuatro semanas del mes de agosto y las primeras tres semanas de septiembre. Los resultados se muestran en la tabla .

	Planta (tratamiento)				
Semana	**1**	**2**	**3**	**4**	$\overline{X}_i$
1	42.7	38.3	42.9	30.1 .	38.5
2	47.3	35.1	38.2	37.5	39.5
3	57.3	42.7	49.9	47.8	49.4
4	63.1	58.2	59.3	53.9	58.6
5	49.2	32.7	45.7	33.8	40.4
6	51.2	30.1	48.3	38.7	42.1
7	48.0	31.1	45.2	39.7	41.0
$\overline{X}_j$	51.3	38.3	47.1	40.2	
$\overline{\overline{X}} = 44.23$					

Melvin realiza un análisis de varianza a una vía y encuentra una diferencia significativa en los niveles de producción promedio. Sin embargo, antes de presentar este reporte a la alta gerencia, Melvin se da cuenta de algo importante: las siete semanas no se seleccionaron aleatoriamente para cada planta. Los datos para las mismas siete semanas se utilizaron para todas las cuatro plantas. Quizá debería hacer bloques en las semanas para eliminar toda variación debida al período. Ya que las mismas semanas se registraron para cada planta, es posible el bloqueo en semanas.

Solución

SCT y SCTR se calculan de la misma forma que con el problema resuelto 1 y se halla que son 2,276.1 y 761.4, respectivamente. Además:

$$\begin{aligned} SCBL &= \Sigma c_i(\overline{X}_i - \overline{\overline{X}})^2 \\ &= 4(38.5 - 44.23)^2 + 4(39.5 - 44.23)^2 \\ &\quad + \cdots + 4(41 - 44.23)^2 \\ &= 1,276.6 \end{aligned}$$

$$SCE = SCT - SCTR - SCBL$$
$$= 238.1$$
$$CMTR = \frac{SCTR}{c-1}$$
$$= \frac{761.4}{4-1}$$
$$= 253.8$$
$$CMBL = \frac{SCBL}{r-1}$$
$$= \frac{1{,}276.6}{7-1}$$
$$= 212.8$$
$$CME = \frac{SCE}{(r-1)(c-1)}$$
$$= \frac{238.1}{(7-1)(4-1)}$$
$$= 13.2$$

Estos cálculos se resumen en la tabla de análisis de varianza a dos vías:

Fuente de variación	SC	g.l.	CM	Valor F
Entre muestras (tratamiento)	761.4	3	253.8	19.23
Entre bloques	1,276.6	6	212.8	16.12
Dentro de muestras (error)	238.1	18	13.2	
Variación total	2,276.1	27		

Primero, Melvin debe determinar si el bloqueo sobre las semanas es efectivo. Él fija α al 5%. El valor F para los bloques es *CMBL/CME*, y debido a que *CMBL* tiene $r - 1 = 6$ g.l. y CME tiene $(r - c)(c - 1) = 18$ g.l., $F_{0.05,6,18} = 2.66$ se halla que es el valor crítico de F. Ya que $F = 16.12 > 2.66$, Melvin concluye que la producción promedio entre semanas es diferente. Por consiguiente, el bloqueo es necesario para corregir la variación de una semana a la otra. Melvin debería continuar con esta prueba a dos vías.

Él puede probar ahora la hipótesis primaria sobre los niveles de producción promedio de la planta

$$H_0\text{: } \mu_1 = \mu_2 = \mu_3 = \mu_4$$
$$H_A\text{: No todas las medias son iguales}$$

El valor F para tratamientos tiene $c - 1 = 3$ y $(r - 1)(c - 1) = 18$ g.l. $F_{0.05,3,18} = 3.16 < 19.23$. Se rechaza la hipótesis, y Melvin concluye que hay alguna diferencia en los niveles de producción promedio. Ahora puede utilizar el *DMS* para determinar cuáles son diferentes.