12

Análisis de la varianza

Objetivos de aprendizaje

Al concluir el capítulo, será capaz de:

OA1 Enumerar las características de la distribución *F* y localizar valores en una tabla *F*.

OA2 Realizar una prueba de hipótesis para determinar si las varianzas de dos poblaciones son iguales.

OA3 Describir el enfoque ANOVA para probar diferencias en medias muestrales.

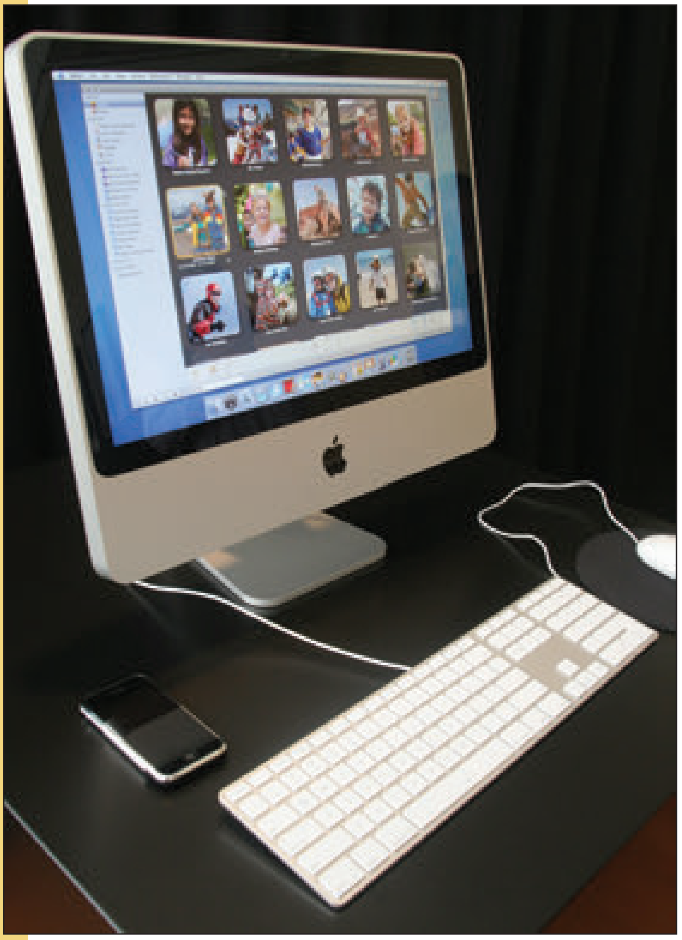
OA4 Organizar datos en una tabla ANOVA para su análisis.

OA5 Realizar una prueba de hipótesis entre tres o más medias de tratamiento y describir los resultados.

OA6 Desarrollar los intervalos de confianza de la diferencia entre medias de tratamiento e interpretar los resultados.

OA7 Realizar una prueba de hipótesis entre medias de tratamiento con una variable de bloqueo.

OA8 Realizar una ANOVA de dos vías con interacción y describir los resultados.



Un fabricante de computadoras está a punto de presentar una nueva computadora personal más rápida. Sin duda, la máquina nueva es más rápida, pero las pruebas iniciales indican que el tiempo de procesamiento varía más, variación que depende del programa que se ejecute, y de la cantidad de datos de entrada y salida. Una muestra de 16 corridas de la computadora, con diversos trabajos de producción, reveló que la desviación estándar del tiempo de procesamiento de la máquina nueva fue de 22 (centésimas de segundo) y de 12 (centésimas de segundo) la del modelo actual. Con un nivel de significancia de 0.05, ¿puede concluir que el tiempo de procesamiento de la máquina nueva varía más? (ejercicio 24, objetivo 2).

12.1 Introducción

En este capítulo se continúa el análisis de las pruebas de hipótesis. Recuerde que en los capítulos 10 y 11 estudió la teoría general de las pruebas de hipótesis. Se analizó el caso en que se seleccionó una muestra de una población. Se utilizó la distribución *z* (la distribución normal estándar) o la distribución *t* para determinar si era razonable concluir que la media poblacional era igual a un valor específico. Se probó si dos medias poblacionales eran iguales. También se realizaron pruebas de una y dos muestras de las proporciones de las poblaciones, con la distribución normal estándar como la distribución del estadístico de prueba. En este capítulo se amplía la idea de pruebas de hipótesis. Se describe una prueba para varianzas y, después, una prueba que compara de forma simultánea varias medias para determinar si provienen de poblaciones iguales.

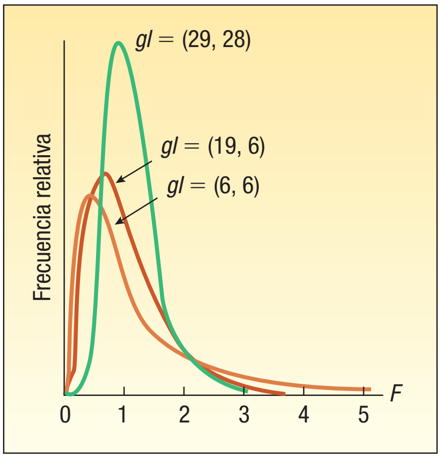
12.2 La distribución *F*

La distribución de probabilidad que se emplea en este capítulo es la distribución *F,* la cual debe su nombre a sir Ronald Fisher, uno de los pioneros de la estadística actual. Esta distribución de probabilidad sirve como la distribución del estadístico de prueba en varias situaciones. Con ella se pone a prueba si dos muestras provienen de poblaciones que tienen varianzas iguales, y también se aplica cuando se desea comparar varias medias poblacionales en forma simultánea. La comparación simultánea de varias medias poblacionales se denomina **análisis de la varianza (ANOVA)**. En las dos situaciones, las poblaciones deben seguir una distribución normal, y los datos deben ser al menos de escala de intervalos.

¿Cuáles son las características de la distribución *F*?

OA1 Enumerar las características de la distribución *F* y localizar valores en una tabla *F*.

1. **Existe una familia de distribuciones *F*.** Cada miembro de la familia se determina mediante dos parámetros: los grados de libertad del numerador y los grados de libertad del denominador. La forma de la distribución se ilustra en la siguiente gráfica. Hay una distribución *F* de la combinación de 29 grados de libertad del numerador (*gl*) y los 28 grados de libertad del denominador. Existe otra distribución *F* de los 19 grados en el numerador y los 6 grados de libertad del denominador. La distribución final que se muestra tiene 6 grados de libertad en el numerador y 6 grados de libertad en el denominador. Los grados de libertad se describen más adelante en este capítulo. Observe que la forma de las curvas cambia cuando varían los grados de libertad.



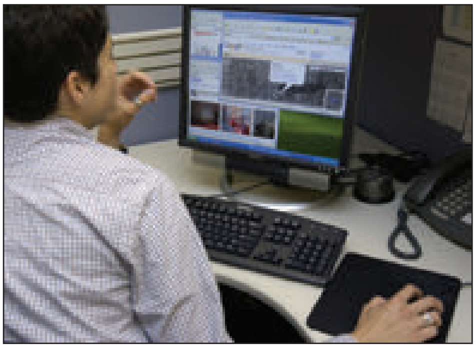
2. **La distribución *F* es continua.** Esto significa que se supone un número infinito de valores entre cero y el infinito positivo.
3. **La distribución *F* no puede ser negativa.** El menor valor que *F* puede tomar es 0.

4. **Tiene sesgo positivo.** La cola larga de la distribución es hacia el lado derecho. Cuando el número de grados de libertad aumenta, tanto en el numerador como en el denominador, la distribución se aproxima a ser normal.
5. **Es asintótica.** Cuando los valores de X aumentan, la curva F se aproxima al eje X pero nunca lo toca. Este caso es similar al comportamiento de la distribución de probabilidad normal, descrito en el capítulo 7.

12.3 Comparación de dos varianzas poblacionales

La primera aplicación de la distribución F ocurre cuando se pone a prueba la hipótesis de que la varianza de una población normal es igual a la varianza de otra población normal. En los siguientes ejemplos se muestra el uso de la prueba:

- Dos máquinas esquiladoras de la marca Barth se calibran para producir barras de acero con la misma longitud. Por lo tanto, las barras deberán tener la misma longitud media. Se desea tener la seguridad de que además de la misma longitud media también tengan una variación similar.



- El índice de rendimiento medio de los dos tipos de acciones comunes puede ser el mismo, pero quizá varíe más el índice de rendimiento de un tipo que el otro. Una muestra de 10 acciones relacionadas con la tecnología y 10 acciones de compañías de servicios presentan el mismo índice de rendimiento medio, pero es probable que varíen más las acciones vinculadas a la tecnología.
- Un estudio del departamento de marketing de un periódico importante reveló que los hombres y las mujeres pasan cerca de la misma cantidad de tiempo por día navegando por la red. Sin embargo, el mismo reporte indica que la variación del tiempo pasado por día por los hombres casi duplicaba al de las mujeres.

La distribución F también sirve para probar suposiciones de algunas pruebas estadísticas. Recuerde que en el capítulo anterior se utilizó la prueba t para investigar si las medias de dos poblaciones independientes eran diferentes. Para emplear esa prueba, algunas veces se supone que las varianzas de dos poblaciones normales son iguales. Vea la lista de suposiciones en la sección 11.4, página 384. La distribución F proporciona un medio para realizar una prueba considerando las varianzas de dos poblaciones normales.

OA2 Realizar una prueba de hipótesis para determinar si las varianzas de dos poblaciones son iguales.

Sin importar si se desea determinar si una población varía más que otra o validar una suposición de una prueba estadística, primero se formula la hipótesis nula. La hipótesis nula es que la varianza de una población normal, σ_1^2, es igual a la varianza de otra población normal, σ_2^2. La hipótesis alternativa podría ser que las varianzas difieren. En este caso, las hipótesis nula y alternativa son:

$$H_0: \sigma_1^2 = \sigma_2^2$$
$$H_1: \sigma_1^2 \neq \sigma_2^2$$

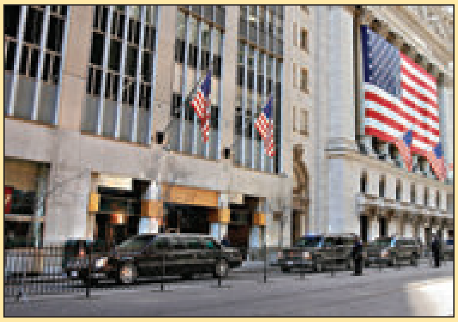
Para realizar la prueba, se selecciona una muestra aleatoria de n_1 observaciones de una población y una muestra aleatoria de n_2 observaciones de la segunda población. El estadístico de prueba se define como sigue.

ESTADÍSTICO DE PRUEBA PARA COMPARAR DOS VARIANZAS

$$F = \frac{s_1^2}{s_2^2} \qquad \textbf{(12-1)}$$

Los términos s_1^2 y s_2^2 son las varianzas muestrales respectivas. Si la hipótesis nula es verdadera, el estadístico de prueba sigue la distribución F con $n_1 - 1$ y $n_2 - 1$ grados de libertad. A fin de reducir el tamaño de la tabla de valores críticos, la varianza *más grande* de la muestra se coloca en el numerador; de aquí, la razón F que se indica en la tabla siempre es mayor que 1.00. Así, el valor crítico de la cola derecha es el único que se requiere. El valor crítico de F de una prueba de dos colas se determina dividiendo el nivel de significancia entre dos ($\alpha/2$) y después se consultan los grados de libertad apropiados en el apéndice B.4. Un ejemplo servirá de ilustración.

Ejemplo



Lammers Limos ofrece servicio de transporte en limusina del ayuntamiento de Toledo, Ohio, al aeropuerto metropolitano de Detroit. Sean Lammers, presidente de la compañía, considera dos rutas. Una por la carretera 25 y la otra por la autopista I-75. Lammers desea estudiar el tiempo que tardaría en conducir al aeropuerto por cada una de las rutas y luego comparar los resultados. Recopiló los siguientes datos muestrales, reportados en minutos. Usando el nivel de significancia de 0.10, ¿hay alguna diferencia entre las variaciones de los tiempos de manejo por las dos rutas?

Carretera 25	Autopista 1-75
52	59
67	60
56	61
45	51
70	56
54	63
64	57
	65

Solución

Los tiempos de manejo medios por las dos rutas son casi iguales. El tiempo medio es de 58.29 minutos por la carretera 25 y de 59.0 minutos por la autopista I-75. Sin embargo, al evaluar los tiempos de recorrido, Lammers también está interesado en la variación de ellos. El primer paso es calcular las dos varianzas muestrales. Se empleará la fórmula (3-11) para determinar la desviación estándar de cada muestra; para obtener la varianza muestral se eleva al cuadrado la desviación estándar.

Carretera 25

$$\bar{X} = \frac{\Sigma X}{n} = \frac{408}{7} = 58.29 \qquad s = \sqrt{\frac{\Sigma(X - \bar{X})^2}{n - 1}} = \sqrt{\frac{485.43}{7 - 1}} = 8.9947$$

Autopista I-75

$$\bar{X} = \frac{\Sigma X}{n} = \frac{472}{8} = 59.00 \qquad s = \sqrt{\frac{\Sigma(X - \bar{X})^2}{n - 1}} = \sqrt{\frac{134}{8 - 1}} = 4.3753$$

Según la medición de la desviación estándar, hay más variación en la carretera 25 que en la autopista I-75. Esto coincide con su conocimiento de las dos rutas; la ruta por la carretera 25 tiene más semáforos, en tanto que la autopista I-75 es de acceso limitado. Sin embargo, la ruta por la autopista I-75 es varias millas más larga. Es importante que el servicio que ofrece sea tanto puntual como consistente, por lo que decide realizar una prueba estadística para determinar si en realidad existe una diferencia entre las variaciones de las dos rutas.

Empleará el procedimiento habitual de la prueba de hipótesis de cinco pasos.

Paso 1: Inicia por formular las hipótesis nula y alternativa. La prueba es de dos colas debido a que se busca una diferencia entre las variaciones de las dos rutas. *No* se trata de demostrar que el tiempo que se emplea varía más por una ruta que por la otra.

$$H_0: \sigma_1^2 = \sigma_2^2$$
$$H_1: \sigma_1^2 \neq \sigma_2^2$$

Paso 2: Selecciona el nivel de significancia de 0.10.

Paso 3: El estadístico de prueba apropiado sigue la distribución *F*.

Paso 4: Obtiene el valor crítico del apéndice B.4, del cual se reproduce una parte como tabla 12-1. Puesto que conduce una prueba de dos colas, el nivel de significancia en la tabla es 0.05, determinado mediante $\alpha/2 = 0.10/2 = 0.05$. Hay $n_1 - 1 = 7 - 1 = 6$ grados de libertad en el numerador, y $n_2 - 1 = 8 - 1 = 7$ grados de libertad en el denominador. Para encontrar el valor crítico, recorre en forma horizontal la parte superior de la tabla *F* (tabla 12-1 o apéndice B.4) del nivel de significancia 0.05 con 6 grados de libertad en el numerador. Después va hacia abajo por esa columna hasta el valor crítico opuesto a 7 grados de libertad en el denominador. El valor crítico es 3.87. Por lo tanto, la regla de decisión es: rechazar la hipótesis si la razón de las varianzas muestrales es mayor que 3.87.

TABLA 12-1 **Valores críticos de la distribución *F*, $\alpha = 0.05$**

Grados de libertad del denominador	Grados de libertad del numerador			
	5	6	7	8
1	230	234	237	239
2	19.3	19.3	19.4	19.4
3	9.01	8.94	8.89	8.85
4	6.26	6.16	6.09	6.04
5	5.05	4.95	4.88	4.82
6	4.39	4.28	4.21	4.15
7	3.97	3.87	3.79	3.73
8	3.69	3.58	3.50	3.44
9	3.48	3.37	3.29	3.23
10	3.33	3.22	3.14	3.07

Paso 3: Por último debe tomar la razón de las dos varianzas muestrales, determinar el valor del estadístico de prueba y tomar una decisión respecto de la hipótesis nula. Observe que la fórmula (12-1) se refiere a las *varianzas* muestrales, pero se calcularon las *desviaciones estándares* de las muestras, las cuales se deben elevar al cuadrado para determinar las varianzas.

$$F = \frac{s_1^2}{s_2^2} = \frac{(8.9947)^2}{(4.3753)^2} = 4.23$$

La decisión es rechazar la hipótesis nula, debido a que el valor *F* calculado (4.23) es mayor que el valor crítico (3.87). Se concluye que hay una diferencia entre las variaciones de los tiempos de recorrido por las dos rutas.

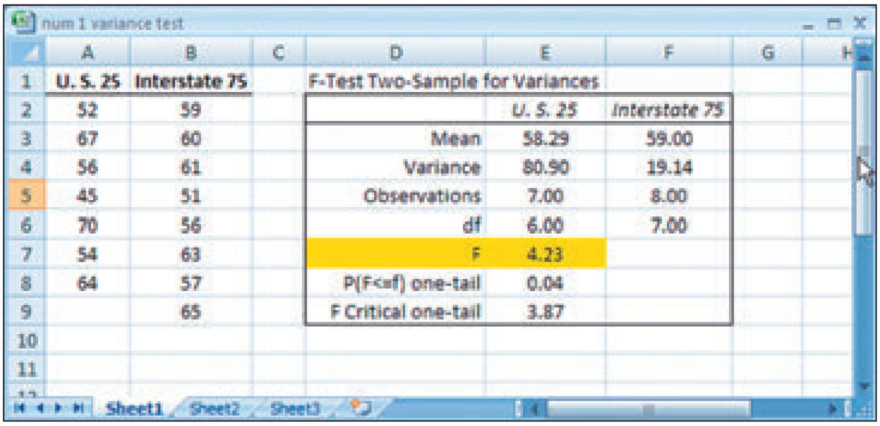
Como se hizo notar, la práctica habitual es determinar la razón F poniendo la mayor de las dos varianzas muestrales en el numerador, lo cual hará que la razón F sea al menos 1.00 y permitirá utilizar siempre la cola derecha de la distribución F para evitar la necesidad de requerir tablas F más extensas.

Respecto de las pruebas de una cola surge una duda lógica. Por ejemplo, suponga que en el ejemplo anterior sospecha que la varianza de los tiempos en la carretera 25 es *mayor* que la varianza de los tiempos por la autopista I-75. Las hipótesis nula y alternativa deberán ser de la siguiente forma:

$$H_0: \sigma_1^2 \leq \sigma_2^2$$
$$H_1: \sigma_1^2 > \sigma_2^2$$

El estadístico de prueba se calcula como s_1^2/s_2^2. Observe que se designó población 1 a la que se sospecha que tiene la varianza mayor. Por lo tanto, s_1^2 aparece en el numerador. La razón F será mayor que 1.00, por lo que se puede utilizar la cola superior de la distribución F. Con estas condiciones, no es necesario dividir el nivel de significancia a la mitad. Como en el apéndice B.4 sólo se dan niveles de significancia de 0.05 y 0.10, estamos restringidos a estos niveles en el caso de pruebas de una cola y con 0.10 y 0.02 en el de pruebas de dos colas, a menos que se consulte una tabla más completa o se utilice software estadístico para calcular el estadístico F.

El programa Excel tiene un procedimiento para realizar una prueba de varianzas. A continuación se presenta la captura de pantalla. El valor calculado de F es el mismo que se determinó con la fórmula (12-1).



num 1 variance test

	A	B	C	D	E	F
1	U. S. 25	Interstate 75		F-Test Two-Sample for Variances		
2	52	59			U. S. 25	Interstate 75
3	67	60		Mean	58.29	59.00
4	56	61		Variance	80.90	19.14
5	45	51		Observations	7.00	8.00
6	70	56		df	6.00	7.00
7	54	63		F	4.23	
8	64	57		P(F<=f) one-tail	0.04	
9		65		F Critical one-tail	3.87	

Sheet1 Sheet2 Sheet3

Autoevaluación 12-1

Steele Electric Products, Inc., ensambla componentes eléctricos para teléfonos celulares. Durante los últimos 10 días, Mark Nagy ha promediado 9 productos rechazados, con una desviación estándar de 2 rechazos por día. Debbie Richmond promedió 8.5 productos rechazados, con una desviación estándar de 1.5 rechazos durante el mismo periodo. Con un nivel de significancia de 0.05, ¿podría concluir que hay más variación en el número de productos rechazados por día de Mark?

Ejercicios

connect

1. ¿Cuál es el valor crítico F de una muestra de seis observaciones en el numerador y cuatro en el denominador? Utilice una prueba de dos colas y el nivel de significancia de 0.10.
2. ¿Cuál es el valor crítico F de una muestra de cuatro observaciones en el numerador y siete en el denominador? Utilice una prueba de una cola y el nivel de significancia de 0.01.

3. Se dan las siguientes hipótesis.

$$H_0: \sigma_1^2 = \sigma_2^2$$
$$H_1: \sigma_1^2 \neq \sigma_2^2$$

En una muestra aleatoria de ocho observaciones de la primera población resultó una desviación estándar de 10. En una muestra aleatoria de seis observaciones de la segunda población resultó una desviación estándar de 7. A un nivel de significancia de 0.02, ¿hay alguna diferencia entre las variaciones de las dos poblaciones?

4. Se dan las siguientes hipótesis.

$$H_0: \sigma_1^2 \leq \sigma_2^2$$
$$H_1: \sigma_1^2 > \sigma_2^2$$

En una muestra aleatoria de cinco observaciones de la primera población resultó una desviación estándar de 12. Una muestra aleatoria de siete observaciones de la segunda población reveló una desviación estándar de 7. A un nivel de significancia de 0.01, ¿varía más la primera población?

5. Arbitron Media Research, Inc., realiza un estudio sobre los hábitos de escuchar iPod de hombres y mujeres. Una parte del estudio incluyó el tiempo medio de escucha. Se descubrió que el tiempo medio de escucha de los hombres era de 35 minutos por día. La desviación estándar de la muestra de los 10 hombres estudiados fue de 10 minutos por día. El tiempo medio de escucha de las 12 mujeres estudiadas también fue de 35 minutos, pero la desviación estándar muestral fue de 12 minutos. A un nivel de significancia de 0.10, ¿puede concluir que hay una diferencia entre las variaciones de los tiempos de escucha de los hombres y las mujeres?
6. Un corredor de bolsa de Critical Securities reportó que la tasa de rendimiento media de una muestra de 10 acciones de la industria petrolera era de 12.6%, con una desviación estándar de 3.9%. La tasa de rendimiento media de una muestra de 8 acciones de compañías de servicios fue de 10.9%, con una desviación estándar de 3.5%. A un nivel de significancia de 0.05, ¿puede concluir que varían más las acciones de la industria petrolera?

12.4 Suposiciones en el análisis de la varianza (ANOVA)

Otro uso de la distribución *F* es el análisis de la técnica de la varianza (ANOVA), en la cual se comparan tres o más medias poblacionales para determinar si pueden ser iguales. Para emplear ANOVA, se supone lo siguiente:

1. Las poblaciones siguen la distribución normal.
2. Las poblaciones tienen desviaciones estándares iguales (σ).
3. Las poblaciones son independientes.

Cuando se cumplen estas condiciones, *F* se emplea como la distribución del estadístico de prueba.

¿Por qué es necesario estudiar ANOVA? ¿Por qué no sólo se emplea la prueba de las diferencias entre medias poblacionales, como se analizó en el capítulo anterior? Se puede comparar dos medias poblacionales a la vez. La razón más importante es la acumulación indeseable del error tipo I. Para ampliar la explicación, suponga cuatro métodos distintos (A, B, C y D) para capacitar personal para ser bomberos. La asignación de cada uno de los 40 prospectos del grupo de este año es aleatoria en cada uno de los cuatro métodos. Al final del programa de capacitación, a los cuatro grupos se les administra una prueba común para medir la comprensión de las técnicas contra incendios. La pregunta es: ¿existe una diferencia entre las calificaciones medias del examen de los cuatro grupos? La respuesta a esta pregunta permitirá comparar los cuatro métodos de capacitación.

El uso de la distribución *t* conduce a la acumulación del error tipo I.

Si emplea la distribución *t* para comparar las cuatro medias poblacionales, tendría que efectuar seis pruebas *t* distintas. Es decir, necesitaría comparar las calificaciones medias de los cuatro métodos como sigue: A contra B, A contra C, A contra D, B contra C, B contra D y C contra D. Si el nivel de significancia es de 0.05, la probabilidad de una decisión estadística correcta es de 0.95, calculada de 1 – 0.05. Como se realizaron seis pruebas separadas (independientes), la probabilidad de que *no* se tome una decisión incorrecta debido al error de muestreo en cualquiera de las seis pruebas independientes es:

$$P(\text{todas correctas}) = (0.95)(0.95)(0.95)(0.95)(0.95)(0.95) = 0.735$$

Para encontrar la probabilidad de que al menos tenga un error debido al muestreo, reste este resultado a 1. Por lo tanto, la probabilidad de al menos una decisión incorrecta debida al muestreo es de 1 – 0.735 = 0.265. En resumen, si realiza seis pruebas independientes con la distribución *t*, la posibilidad de rechazar una hipótesis nula verdadera debido al error de muestreo se incrementa de 0.05 a un nivel insatisfactorio de 0.265. Es obvio que necesita un mejor método que realizar seis pruebas *t*. ANOVA le permite comparar las medias de tratamiento de forma simultánea y evitar la acumulación del error de tipo I.

OA3 Describir el enfoque ANOVA para probar diferencias entre medias muestrales.

ANOVA se desarrolló para aplicaciones en agricultura, y aún se emplean muchos de los términos relacionados con ese contexto. En particular, con el término *tratamiento* se identifican las diferentes poblaciones que se examinan. Por ejemplo, el tratamiento se refiere a cómo se trató una extensión de terreno con un tipo particular de fertilizante. La siguiente ilustración aclarará el término *tratamiento* y mostrará la aplicación de ANOVA.

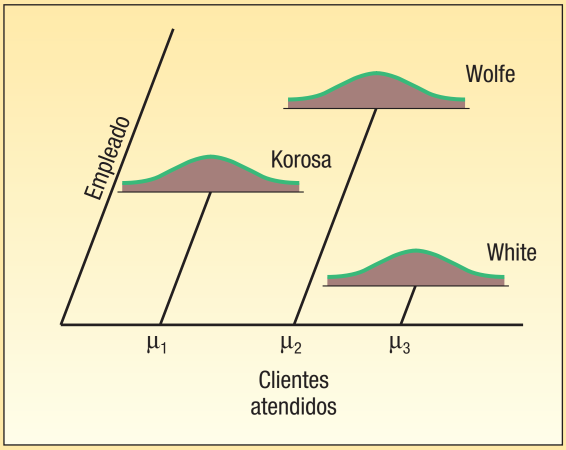
Ejemplo

Joyce Kuhlman es gerente de un centro financiero regional y desea comparar la productividad, medida por el número de clientes atendidos, de tres empleados. Selecciona cuatro días en forma aleatoria y registra el número de clientes que atendió cada empleado. Los resultados son:

Wolfe	White	Korosa
55	66	47
54	76	51
59	67	46
56	71	48

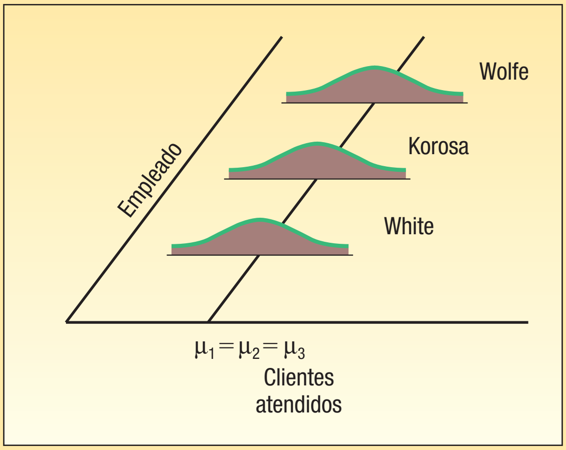
Solución

¿Hay alguna diferencia en el número medio de clientes atendidos? En la gráfica 12-1 se ilustra cómo pueden aparecer las poblaciones si hubiera una diferencia en las medias del tratamiento. Observe que las poblaciones siguen la distribución normal y la variación en cada población es la misma. Sin embargo, las medias *no* son iguales.



GRÁFICA 12-1 Caso en el que las medias del tratamiento son diferentes

Suponga que las poblaciones son iguales. Es decir, que no hay una diferencia entre las medias (tratamiento). Esta igualdad, que se muestra en la gráfica 12-2, indica que las medias poblacionales son iguales. Observe de nuevo que las poblaciones siguen la distribución normal, y que la variación en cada una de las poblaciones es la misma.



GRÁFICA 12-2 Caso en el que las medias del tratamiento son iguales

12.5 La prueba ANOVA

OA5 Realizar una prueba de hipótesis entre tres o más medias de tratamiento y describir los resultados.

¿Cómo funciona la prueba ANOVA? Recuerde que se desea determinar si varias medias muestrales provienen de una sola población o de poblaciones con medias diferentes. En realidad, estas medias muestrales se comparan mediante sus varianzas. Para explicar esto, recuerde que en la página 416 se enumeraron las suposiciones que requiere ANOVA. Una de estas suposiciones fue que las desviaciones estándares de las diversas poblaciones normales tenían que ser las mismas. Se aprovecha este requisito en la prueba ANOVA. La estrategia es estimar la varianza de la población (desviación estándar al cuadrado) de dos formas para después determinar la razón de dichas estimaciones. Si esta razón es aproximadamente 1, entonces por lógica las dos estimaciones son iguales, y se concluye que las medias poblacionales no son iguales. La distribución F sirve como un árbitro para indicar en qué instancia la razón de las varianzas muestrales es mucho mayor que 1 para haber ocurrido por casualidad.

Consulte el ejemplo del centro financiero en la sección anterior. El gerente desea determinar si hay una diferencia entre los números medios de clientes atendidos. Para iniciar, determine la media global de las 12 observaciones. Ésta es de 58, calculada de $(55 + 54 + \cdots + 48)/12$. Después, en cada una de las 12 observaciones encuentre la diferencia entre el valor particular y la media global. Cada una de estas diferencias se eleva al cuadrado y estos cuadrados se suman. Este término se denomina **variación total**.

VARIACIÓN TOTAL Suma de las diferencias entre cada observación y la media global elevadas al cuadrado.

En nuestro ejemplo, la variación total es de 1 082, determinada por $(55 - 58)^2 + (54 - 58)^2 + \cdots + (48 - 58)^2$.

Luego se divide esta variación total en dos componentes: la que se debe a los **tratamientos** y la que es **aleatoria**. Para encontrar estas dos componentes, se determina la media de cada tratamiento. La primera fuente de variación se debe a los tratamientos.

VARIACIÓN DE TRATAMIENTO Suma de las diferencias entre la media de cada tratamiento y la media total o global elevadas al cuadrado.

En el ejemplo, la variación debida a los tratamientos es la suma de las diferencias al cuadrado entre la media de cada empleado y la media global. Este término es 992. Para calcularlo, primero se encuentra la media de cada uno de los tres tratamientos. La media de Wolfe es 56, determinada por (55 + 54 + 59 + 56)/4. Las otras medias son 70 y 48, respectivamente. La suma de los cuadrados debida a los tratamientos es:

$$(56 - 58)^2 + (56 - 58)^2 + \cdots + (48 - 58)^2 = 4(56 - 58)^2 + 4(70 - 58)^2 + 4(48 - 58)^2$$
$$= 992$$

Si existe una variación considerable entre las medias de los tratamientos, es lógico que este término sea grande. Si las medias son similares, este término será un valor bajo. El valor más bajo posible es cero. Esto ocurrirá cuando todas las medias de los tratamientos sean iguales.

A la otra fuente de variación se le conoce como componente **aleatoria**, o componente de error.

VARIACIÓN ALEATORIA Suma de las diferencias entre cada observación y su media de tratamiento elevadas al cuadrado.

En el ejemplo, este término es la suma de las diferencias al cuadrado entre cada valor y la media de ese empleado en particular. La variación de error es 90.

$$(55 - 56)^2 + (54 - 56)^2 + \cdots + (48 - 48)^2 = 90$$

El estadístico de prueba, que es la razón de las dos estimaciones de la varianza poblacional, se determina a partir de la siguiente ecuación:

$$F = \frac{\text{Estimación de la varianza poblacional basada en las diferencias entre las medias muestrales}}{\text{Estimación de la varianza poblacional basada en la variación dentro de la muestra}}$$

La primera estimación de la varianza poblacional parte de los tratamientos, es decir, de la diferencia *entre* las medias. Éste es 992/2. ¿Por qué se dividió entre 2? Recuerde del capítulo 3 que, para encontrar una varianza muestral [vea la fórmula (3-11)], se divide entre el número de observaciones menos uno. En este caso hay tres tratamientos, por lo que se divide entre 2. La primera estimación de la varianza poblacional es 992/2.

La estimación de la varianza *dentro* de los tratamientos es la variación aleatoria dividia entre el número total de observaciones menos el número de tratamiento. Es decir 90/(12 – 3). De aquí, la segunda estimación de la varianza poblacional es 90/9. En realidad es una generalización de la fórmula (11-5), en la cual se agruparon las varianzas muestrales de dos poblaciones.

El paso final es tomar la razón de estas dos estimaciones.

$$F = \frac{992/2}{90/9} = 49.6$$

Como esta razón es muy distinta a 1, se concluye que las medias de los tratamientos no son iguales. Hay una diferencia entre los números medios de clientes atendidos por cada uno de los tres empleados.

A continuación se presenta otro ejemplo, el cual trata de muestras de tamaños diferentes.

Ejemplo

Desde hace algún tiempo las aerolíneas han reducido sus servicios, como alimentos y bocadillos durante sus vuelos, y empezaron a cobrar un precio adicional por algunos de ellos, como llevar sobrepeso de equipaje, cambios de vuelo de último momento y por mascotas que viajan en la cabina. Sin embargo, aún están muy preocupadas por el servicio que ofrecen. Hace poco un grupo de cuatro aerolíneas contrató a Brunner Marketing Research, Inc., para encuestar a sus pasajeros sobre la adquisición de boletos, abordaje, servicio durante el vuelo, manejo del equipaje, comunicación del piloto, etc. Hicieron 25 preguntas con diversas respuestas posibles: excelente, bueno, regular o deficiente. Una respuesta de excelente tiene una calificación de 4, bueno 3, regular 2 y deficiente 1. Estas respuestas se sumaron, de modo que la calificación final fue una indicación de la satisfacción con el vuelo. Entre mayor la calificación, mayor el nivel de satisfacción con el servicio. La calificación mayor posible fue 100.

Brunner seleccionó y estudió al azar pasajeros de las cuatro aerolíneas. A continuación se muestra la información. ¿Hay alguna diferencia entre los niveles de satisfacción medios con respecto a las cuatro aerolíneas? Use el nivel de significancia de 0.01.

Northern	WTA	Pocono	Branson
94	75	70	68
90	68	73	70
85	77	76	72
80	83	78	65
	88	80	74
		68	65
		65	

Solución

Utilice el procedimiento de prueba de hipótesis de cinco pasos.

Paso 1: Formule las hipótesis nula y alternativa. La hipótesis nula es que las calificaciones medias de las cuatro aerolíneas son iguales.

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

La hipótesis alternativa es que no todas las calificaciones medias son iguales.

H_1: No todas las calificaciones medias son iguales.

La hipótesis alternativa también se considera como "al menos dos calificaciones medias no son iguales".

Si no se rechaza la hipótesis nula, se concluye que no hay una diferencia entre las calificaciones medias de las cuatro aerolíneas. Si se rechaza H_0, se concluye que hay una diferencia en al menos un par de calificaciones medias, pero en este punto no se sabe cuál par o cuántos pares difieren.

Paso 2: Seleccione el nivel de significancia. Seleccionó el nivel de significancia de 0.01.

Paso 3: Determine el estadístico de prueba. El estadístico de prueba sigue la distribución F.

Paso 4: Formule la regla de decisión. Para determinar la regla de decisión, necesita el valor crítico. El valor crítico del estadístico F aparece en el apéndice B.4. Los valores críticos del nivel de significancia 0.05 se encuentran en la primera página, y el nivel de significancia de 0.01, en la segunda. Para utilizar esta tabla necesita conocer los grados de libertad del numerador y del denominador. Los grados de libertad del numerador son iguales al número de tratamientos, designado k, menos 1. Los grados de libertad del denominador son el número total de observaciones, n, menos el número de tratamientos. En este ejemplo hay cuatro tratamientos y un total de 22 observaciones.

Grados de libertad del numerador $= k - 1 = 4 - 1 = 3$

Grados de libertad del denominador $= n - k = 22 - 4 = 18$

OA4 Organizar datos en una tabla ANOVA para su análisis.

Consulte el apéndice B.4 y el nivel de significancia de 0.01. Muévase horizontalmente por la parte superior de la página a tres grados de libertad del numerador. Después vaya hacia abajo por esa columna hasta la fila con 18 grados de libertad. El valor en esta intersección es 5.09. Por lo tanto, la regla de decisión es rechazar H_0 si el valor calculado de F es mayor que 5.09.

Paso 5: Seleccione la muestra, realice los cálculos y tome una decisión. Es conveniente resumir los cálculos del estadístico F en una **tabla ANOVA**. El formato de una tabla ANOVA es como sigue. En los paquetes de software estadístico también se emplea este formato.

Tabla ANOVA

Fuente de variación	Suma de cuadrados	Grados de libertad	Media cuadrática	*F*
Tratamientos	SST	$k-1$	$\text{SST}/(k-1) = \text{MST}$	MST/MSE
Error	SSE	$n-k$	$\text{SSE}/(n-k) = \text{MSE}$	
Total	SS total	$n-1$		

Hay tres valores, o suma de cuadrados, para calcular el estadístico de prueba F. Estos valores se determinan al obtener SS total y SSE, después SST mediante una resta. El término SS total es la variación total, SST es la variación debida a los tratamientos, y SSE es la variación dentro de los tratamientos o el error aleatorio.

En general, el proceso se inicia al determinar SST total: la suma de las diferencias elevadas al cuadrado entre cada observación y la media global. La fórmula para determinar SS total es:

$$\text{SS total} = \Sigma(X - \overline{X}_G)^2 \qquad \textbf{(12-2)}$$

donde:

X es cada observación de la muestra.

$\overline{X}_G$ es la media global o total.

En seguida se determina SSE o la suma de los errores elevados al cuadrado: la suma de las diferencias elevadas al cuadrado entre cada observación y su respectiva media de tratamiento. La fórmula para encontrar SSE es:

$$\text{SSE} = \Sigma(X - \overline{X}_c)^2 \qquad \textbf{(12-3)}$$

donde:

$\overline{X}_c$ es la media muestral del tratamiento c.

A continuación se presentan los cálculos detallados de SS total y SSE de este ejemplo. Para determinar los valores de SS total y SSE se comienza por calcular la media global o total. Hay 22 observaciones y el total es 1 664, por lo cual la media total es 75.64.

$$\overline{X}_G = \frac{1\,664}{22} = 75.64$$

	Northern	WTA	Pocono	Branson	Total
	94	75	70	68	
	90	68	73	70	
	85	77	76	72	
	80	83	78	65	
		88	80	74	
			68	65	
			65		
Total de la columna	349	391	510	414	1 664
n	4	5	7	6	22
Media	87.25	78.20	72.86	69.00	75.64

Luego se encuentra la desviación de cada observación a la media total: se elevan al cuadrado estas desviaciones y se suma el resultado de las 22 observaciones. Por ejemplo, el primer pasajero encuestado tenía una calificación de 94, y la media global o total es 75.64. Por lo tanto, $(X - \bar{X}_G) = 94 - 75.64 = 18.36$. En el caso del último pasajero, $(X - \bar{X}_G) = 65 - 75.64 = -10.64$. Los cálculos relativos a los otros pasajeros son:

Northern	WTA	Pocono	Branson
18.36	−0.64	−5.64	−7.64
14.36	−7.64	−2.64	−5.64
9.36	1.36	0.36	−3.64
4.36	7.36	2.36	−10.64
	12.36	4.36	−1.64
		−7.64	−10.64
		−10.64	

Después se eleva al cuadrado cada una de estas diferencias y se suman todos los valores. Así, en el caso del primer pasajero:

$$(X - \bar{X}_G)^2 = (94 - 75.64)^2 = (18.36)^2 = 337.09$$

Por último, se suman todas las diferencias elevadas al cuadrado, como se indica en la fórmula (12-2). El valor SS total es 1 485.10.

	Northern	WTA	Pocono	Branson	Total
	337.09	0.41	31.81	58.37	
	206.21	58.37	6.97	31.81	
	87.61	1.85	0.13	13.25	
	19.01	54.17	5.57	113.21	
		152.77	19.01	2.69	
			58.37	113.21	
			113.21		
Total	649.92	267.57	235.07	332.54	1 485.10

Para calcular el término SSE se encuentra la desviación entre cada observación y su media de tratamiento. En el ejemplo, la media del primer tratamiento (es decir, los pasajeros en Northern Airlines) es 87.25, determinada mediante $\bar{X}_N = 349/4$. El subíndice N se refiere a Northern Airlines.

El primer pasajero calificó a Northern con 94, por lo que $(X - \bar{X}_N) = (94 - 87.25) = 6.75$. El primer pasajero del grupo de TWA respondió con una calificación total de 75, por lo cual $(X - \bar{X}_W) = (75 - 78.20) = -3.2$. El detalle de todos los pasajeros es:

Northern	WTA	Pocono	Branson
6.75	−3.2	−2.86	−1
2.75	−10.2	0.14	1
−2.25	−1.2	3.14	3
−7.25	4.8	5.14	−4
	9.8	7.14	5
		−4.86	−4
		−7.86	

Estadística en acción

¿Alguna vez ha estado esperando que se desocupe un teléfono público y la persona que lo usa pareciera hablar sin parar? Existe evidencia de que la gente habla más por un teléfono público cuando alguien está esperando que lo desocupe. En una encuesta reciente en un centro comercial, los investigadores midieron el tiempo que 56 compradores pasaron hablando por teléfono: 1) cuando estaban solos, 2) cuando una persona estaba usando el teléfono de al lado y 3) cuando una persona estaba usando un teléfono de al lado y alguien esperaba su turno. El estudio, que aplicó la técnica ANOVA de una vía, demostró que el tiempo medio de uso del teléfono era significativamente menor cuando la persona estaba sola.

Cada uno de estos valores se eleva al cuadrado y después se suman las 22 observaciones. Los valores se muestran en la siguiente tabla.

	Northern	WTA	Pocono	Branson	Total
	45.5625	10.24	8.18	1	
	7.5625	104.04	0.02	1	
	5.0625	1.44	9.86	9	
	52.5625	23.04	26.42	16	
		96.04	50.98	25	
			23.62	16	
			61.78		
Total	110.7500	234.80	180.86	68	594.41

Por lo tanto, el valor SSE es 594.41. Es decir, $\Sigma(X - \overline{X}_c)^2 = 594.41$.

Por último, se determina SST, la suma de los cuadrados debida a los tratamientos, con la resta:

$$SST = \text{SS total} - SSE \qquad \textbf{(12-4)}$$

En este ejemplo:

$$SST = \text{SS total} - SSE = 1\,485.10 - 594.41 = 890.69$$

Para determinar el valor calculado de F, consulte la tabla ANOVA. Los grados de libertad del numerador y del denominador son los mismos que en el paso 4 en la página 420, donde se determinó el valor crítico de F. El término **media cuadrática** es otra expresión de la estimación de la varianza. La media cuadrática de tratamientos es SST dividido entre sus grados de libertad. El resultado es la **media cuadrática de tratamientos**, y se escribe MST. Calcule el **error medio cuadrático** de una manera similar. Para ser precisos, divida SSE entre sus grados de libertad. Para completar el proceso y obtener F, divida MST entre MSE.

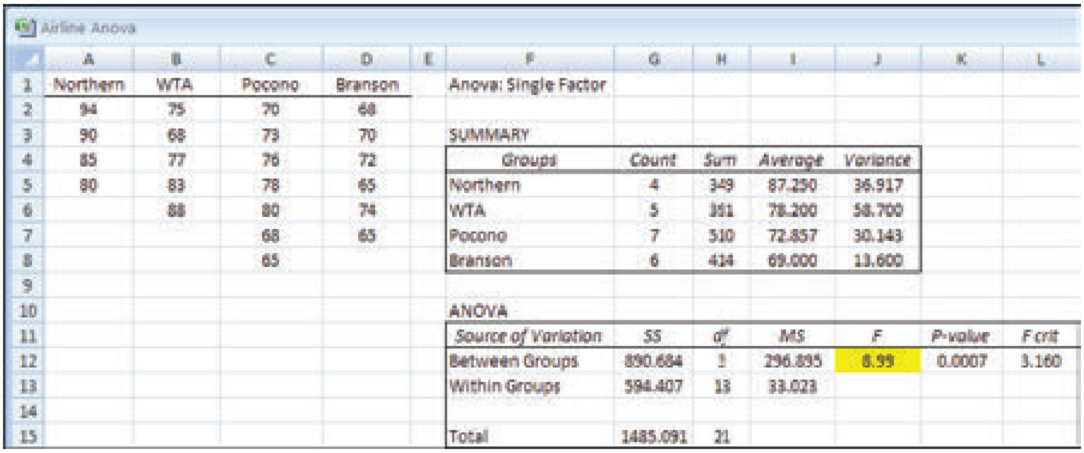
Sustituya los valores particulares de F en una tabla ANOVA y calcule el valor de F, como se muestra a continuación.

Fuente de variación	Suma de cuadrados	Grados de libertad	Media cuadrática	*F*
Tratamientos	890.69	3	296.90	8.99
Error	594.41	18	33.02	
Total	1 485.10	21		

El valor calculado de F es 8.99, mayor que el valor crítico de 5.09, por lo que la hipótesis nula se rechaza. La conclusión es que no todas las medias poblacionales son iguales. Las calificaciones medias de las cuatro aerolíneas no son iguales. Es probable que las calificaciones de los pasajeros se relacionen con una de ellas. En este punto sólo es posible concluir que hay una diferencia entre las medias del tratamiento. No se puede determinar cuáles ni cuántos grupos de tratamientos difieren.

Como se hizo notar en el ejemplo previo, los cálculos son tediosos si la cantidad de observaciones en cada tratamiento es extensa. Hay muchos paquetes de software para generar estos resultados. A continuación se presenta la captura de pantalla de Excel en forma de una tabla ANOVA para el ejemplo anterior, con las calificaciones de aerolíneas y de pasajeros.

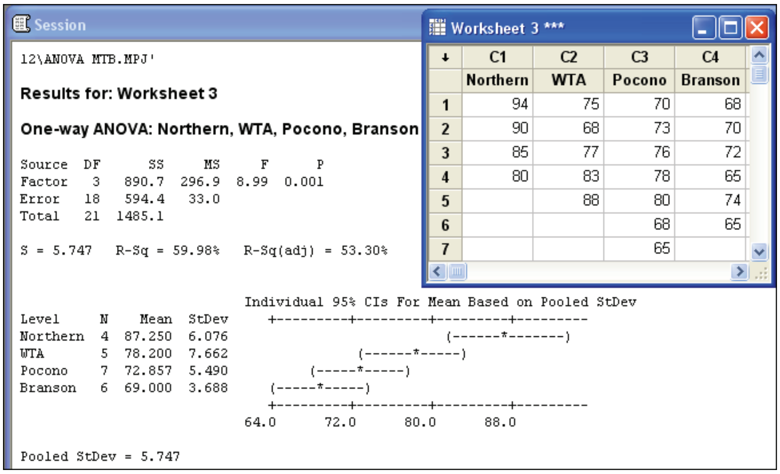
Existen algunas diferencias sutiles entre la captura de pantalla y los cálculos anteriores. Estas diferencias se deben al redondeo.



	A	B	C	D	E	F	G	H	I	J	K	L
1	Northern	WTA	Pocono	Branson		Anova: Single Factor						
2	94	75	70	68								
3	90	68	73	70		SUMMARY						
4	85	77	76	72		Groups	Count	Sum	Average	Variance		
5	80	83	78	65		Northern	4	349	87.250	36.917		
6		88	80	74		WTA	5	391	78.200	58.700		
7			68	65		Pocono	7	510	72.857	30.143		
8			65			Branson	6	414	69.000	13.600		
9												
10						ANOVA						
11						Source of Variation	SS	df	MS	F	P-value	F crit
12						Between Groups	890.684	3	296.895	8.99	0.0007	3.160
13						Within Groups	594.407	18	33.023			
14												
15						Total	1485.091	21				

Observe que en Excel se emplea el término “Between Groups” (Entre grupos) para “Tratamientos”, y “Within Groups” (Dentro de grupos) para “Error”. Sin embargo, tienen el mismo significado. El valor p es 0.0007. Ésta es la probabilidad de determinar un valor del estadístico de prueba de esta magnitud o mayor cuando la hipótesis nula es verdadera. En otras palabras, es la probabilidad de calcular un valor F mayor que 8.99 con 3 grados de libertad en el numerador y 18 grados de libertad en el denominador. Por lo tanto, cuando se rechaza la hipótesis nula en este caso hay una posibilidad muy remota de cometer un error tipo I.

En seguida se presenta la captura de pantalla de Minitab del ejemplo de las calificaciones de los pasajeros de aerolíneas, similar a la captura de pantalla de Excel. La salida también está en la forma de una tabla ANOVA. Además, Minitab proporciona información sobre las diferencias entre medias. Esto se analiza en la siguiente sección.



↓	C1	C2	C3	C4
	Northern	WTA	Pocono	Branson
1	94	75	70	68
2	90	68	73	70
3	85	77	76	72
4	80	83	78	65
5		88	80	74
6			68	65
7			65	

En el sistema Minitab se emplea el término “Factor” en lugar de *tratamiento*, con el mismo significado.

Autoevaluación 12-2

Citrus Clean es un nuevo limpiador multiusos a prueba en el mercado, del cual se han colocado exhibidores en tres lugares distintos dentro de varios supermercados. A continuación se reporta la cantidad de botellas de 12 onzas que se vendieron en cada lugar del supermercado.

Cerca del pan	18	14	19	17
Cerca de la cerveza	12	18	10	16
Cerca de otros limpiadores	26	28	30	32

A un nivel de significancia de 0.05, ¿hay alguna diferencia entre los promedios de botellas que se vendieron en los tres lugares?

a) Formule las hipótesis nula y alternativa.
b) ¿Cuál es la regla de decisión?
c) Calcule los valores de SS total, SST y SSE.
d) Elabore una tabla ANOVA.
e) ¿Cuál es su decisión respecto de la hipótesis nula?

Ejercicios

7. La siguiente es información muestral. Verifique la hipótesis de que las medias de tratamiento son iguales. Utilice el nivel de significancia de 0.05.

Tratamiento 1	Tratamiento 2	Tratamiento 3
8	3	3
6	2	4
10	4	5
9	3	4

a) Formule las hipótesis nula y alternativa.
b) ¿Cuál es la regla de decisión?
c) Calcule los valores SST, SSE y SS total.
d) Elabore una tabla ANOVA.
e) Declare su decisión respecto de la hipótesis nula.

8. La siguiente es información muestral. Verifique la hipótesis con un nivel de significancia de 0.05 de que las medias de tratamiento son iguales.

Tratamiento 1	Tratamiento 2	Tratamiento 3
9	13	10
7	20	9
11	14	15
9	13	14
12		15
10		

a) Formule las hipótesis nula y alternativa.
b) ¿Cuál es la regla de decisión?
c) Calcule SST, SSE y SS total.
d) Elabore una tabla ANOVA.
e) Declare su decisión respecto de la hipótesis nula.

9. Un inversionista en bienes raíces considera invertir en un centro comercial en los suburbios de Atlanta, Georgia, para lo cual evalúa tres terrenos. El ingreso familiar en el área circundante al centro comercial propuesto tiene una importancia particular. Se selecciona una muestra aleatoria de cuatro familias cerca de cada centro comercial propuesto. A continuación se presentan los resultados de la muestra. A un nivel de significancia de 0.05, ¿el inversionista puede concluir que hay una diferencia entre los ingresos medios? Utilice el procedimiento de prueba de hipótesis habitual de cinco pasos.

Área de Southwyck (en miles de dólares)	Franklin Park (en miles de dólares)	Old Orchard (en miles de dólares)
64	74	75
68	71	80
70	69	76
60	70	78

10. La gerente de una compañía de software desea estudiar el número de horas que los directivos de diversas empresas utilizan sus computadoras de escritorio. El gerente seleccionó una muestra de cinco ejecutivos de cada una de tres industrias. A un nivel de significancia de 0.05, ¿puede la gerente concluir que hay una diferencia entre los promedios de horas por semana que se utilizan las computadoras en la industria?

Bancaria	Detallista	De seguros
12	8	10
10	8	8
10	6	6
12	8	8
10	10	10

12.6 Tratamiento e inferencia sobre pares de medias

Suponga que realiza el procedimiento ANOVA y toma la decisión de rechazar la hipótesis nula. Esto permite concluir que no todas las medias de tratamiento son iguales. Algunas veces esta conclusión sería satisfactoria, pero otras se desea conocer cuáles medias de tratamiento difieren. En esta sección se proporcionan los detalles de prueba para saber cuáles medias de tratamiento difieren.

Recuerde que en el ejemplo de Brunner Research respecto de las calificaciones que aplicaron los pasajeros de aerolíneas, había una diferencia entre las medias de tratamiento. Es decir, se rechazó la hipótesis nula y se aceptó la hipótesis alternativa. Si las calificaciones de los pasajeros no difieren, la pregunta es: ¿entre qué grupos difieren las medias de tratamiento?

OA6 Desarrollar intervalos de confianza de la diferencia entre medias de tratamiento e interpretar los resultados.

Se dispone de varios procedimientos para responder esta pregunta. El más simple es emplear intervalos de confianza, es decir, la fórmula (9-2). A partir de la captura de pantalla de la computadora del ejemplo anterior (consulte la página 424), observe que la calificación media muestral de los pasajeros del servicio de la aerolínea Northtern es 87.25, mientras la media muestral de los que califican el servicio de la aerolínea Branson es 69.00. ¿Existe suficiente disparidad para justificar la conclusión de que hay una diferencia significativa entre las calificaciones de satisfacción medias de las dos aerolíneas?

La distribución *t*, descrita en los capítulos 10 y 11, sirve como base de esta prueba. Recuerde que una de las suposiciones de ANOVA es que las varianzas poblacionales de todos los tratamientos son las mismas. Este valor común de la población es el **error medio cuadrá-**

tico, o MSE, y se determina mediante SSE/(n – k). Un intervalo de confianza de la diferencia entre dos poblaciones se obtiene mediante:

INTERVALO DE CONFIANZA DE LA DIFERENCIA ENTRE LAS MEDIAS DE TRATAMIENTO

$$(\bar{X}_1 - \bar{X}_2) \pm t\sqrt{\text{MSE}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)} \qquad \textbf{(12-5)}$$

donde:

$\bar{X}_1$ es la media de la primera muestra.
$\bar{X}_2$ es la media de la segunda muestra.
t se obtiene del apéndice B.2. Los grados de libertad son iguales a n – k.
MSE es el error medio cuadrático que se obtuvo de la tabla ANOVA [SSE/(n – k)].
n_1 es el número de observaciones en la primera muestra.
n_2 es el número de observaciones en la segunda muestra.

¿Cómo se decide si hay una diferencia entre las medias de tratamiento? Si el intervalo de confianza incluye cero, *no* existe diferencia entre ellas. Por ejemplo, si el punto extremo izquierdo del intervalo de confianza tiene signo negativo y el punto extremo derecho tiene signo positivo, el intervalo incluye cero, y las dos medias no difieren. Por lo tanto, si se desarrolla un intervalo de confianza a partir de la fórmula (12-5) y se tiene que la diferencia entre las medias muestrales fue 5.00, es decir, si $\bar{X}_1 - \bar{X}_2 = 5$ y $t\sqrt{\text{MSE}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)} = 12$, el intervalo de confianza variará de -7.00 hasta 17.00. Expresado en símbolos:

$$(\bar{X}_1 - \bar{X}_2) \pm t\sqrt{\text{MSE}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)} = 5.00 \pm 12.00 = -7.00 \text{ hasta } 17.00$$

Observe que en este intervalo se incluye el cero. Por ello, se concluye que no hay una diferencia significativa entre las medias de tratamiento seleccionadas.

Por otro lado, si los puntos extremos del intervalo de confianza tienen el mismo signo, esto indica que las medias de tratamiento difieren. Por ejemplo, si $\bar{X}_1 - \bar{X}_2 = -0.35$ y $t\sqrt{\text{MSE}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)} = 0.25$, el intervalo de confianza variará de -0.60 hasta -0.10. Como -0.60 y -0.10 tienen el mismo signo, ambos negativos, cero no se encuentra en el intervalo y se concluye que estas medias de tratamiento difieren.

Use el ejemplo anterior sobre las aerolíneas para calcular el intervalo de confianza de la diferencia entre las calificaciones medias de los pasajeros de las aerolíneas Northern y Branson. Con un nivel de confianza de 95%, los puntos extremos del intervalo de confianza son 10.46 y 26.04.

$$\begin{aligned}(\bar{X}_A - \bar{X}_{US}) \pm t\sqrt{\text{MSE}\left(\frac{1}{n_A} + \frac{1}{n_{US}}\right)} &= (87.25 - 69.00) \pm 2.101\sqrt{33.0\left(\frac{1}{4} + \frac{1}{6}\right)} \\ &= 18.25 \pm 7.79\end{aligned}$$

donde:

$\bar{X}_A$ es 87.25.
$\bar{X}_{US}$ es 69.00
t es 2.101: del apéndice B.2 con (n – k) = 22 – 4 = 18 grados de libertad.
MSE es 33.0: de la tabla ANOVA con SSE/(n – k) = 594.4/18.
n_E es 4.
n_{US} es 6.

El intervalo de confianza de 95% varía de 10.46 hasta 26.04. Los dos puntos extremos son positivos; de aquí se puede concluir que estas medias de tratamiento difieren de manera significativa. Es decir, los pasajeros de Northern calificaron el servicio en gran medida diferente de los de Branson Airlines.

También es posible obtener resultados aproximados de forma directa a partir de la captura de pantalla de Minitab. A continuación se presenta la parte inferior de la captura de pantalla que se muestra en la página 424. A la izquierda se encuentra el número de observaciones, la media y la desviación estándar de cada tratamiento. Siete pasajeros de Allegheny calificaron su servicio con 72.857, con una desviación estándar de 5.490.

```
                                 Individual 95% CIs For Mean Based on Pooled StDev
Level      N    Mean   StDev     +---------+---------+---------+---------
Northern   4  87.250   6.076                           (------*-------)
WTA        5  78.200   7.662                (------*-----)
Pocono     7  72.857   5.490          (-----*-----)
Branson    6  69.000   3.688     (-----*-----)
                                 +---------+---------+---------+---------
                               64.0      72.0      80.0      88.0
```

A la derecha de la impresión se encuentra un intervalo de confianza cada media de tratamiento. El asterisco (*) indica la ubicación de la media de tratamiento, y la apertura de paréntesis a la izquierda y el cierre de paréntesis a la derecha, los puntos extremos del intervalo de confianza. En los casos donde se superponen los paréntesis, quizá no difieran las medias de tratamiento. Si no hay un área común en los intervalos de confianza, ese par de medias difiere.

Los puntos extremos de un intervalo de confianza de 95% de las calificaciones de los pasajeros de la compañía Pocono son aproximadamente 69 y 77. Los puntos extremos del intervalo de confianza de 95% de la compañía Branson de la calificación media de los pasajeros son aproximadamente 64 y 73. Hay un área común entre estos puntos, por lo cual se concluye que este par de medias no difieren. En otras palabras, no hay una diferencia significativa entre las calificaciones medias de los pasajeros de las aerolíneas Pocono y Branson. La diferencia entre las calificaciones medias se debe a la casualidad.

Hay dos pares de medias que difieren. Las calificaciones medias de los pasajeros de la aerolínea Northern difieren de manera significativa de las calificaciones medias de los pasajeros de las aerolíneas Pocono y Branson. No hay un área común entre estos pares de intervalos de confianza.

Se debe destacar que esta investigación es un proceso que avanza por pasos. El paso inicial es realizar la prueba ANOVA. Sólo si se rechaza la hipótesis nula de que las medias de tratamiento son iguales se deberán analizar las medias de tratamiento individuales.

Autoevaluación 12-3

Los siguientes datos son las colegiaturas por semestre (en miles de dólares) de una muestra de universidades privadas en varias regiones de Estados Unidos. A un nivel de significancia de 0.05, ¿se puede concluir que hay una diferencia entre las colegiaturas medias de las diversas regiones?

Noreste (en miles de dólares)	Sureste (en miles de dólares)	Oeste (en miles de dólares)
10	8	7
11	9	8
12	10	6
10	8	7
12		6

a) Formule las hipótesis nula y alternativa.
b) ¿Cuál es la regla de decisión?
c) Elabore una tabla ANOVA. ¿Cuál es el valor del estadístico de prueba?
d) ¿Cuál es su decisión respecto de la hipótesis nula?
e) ¿Puede existir una diferencia significativa entre la colegiatura media en el noreste en comparación con la del oeste? Si la hay, desarrolle el intervalo de confianza de 95% de esa diferencia.