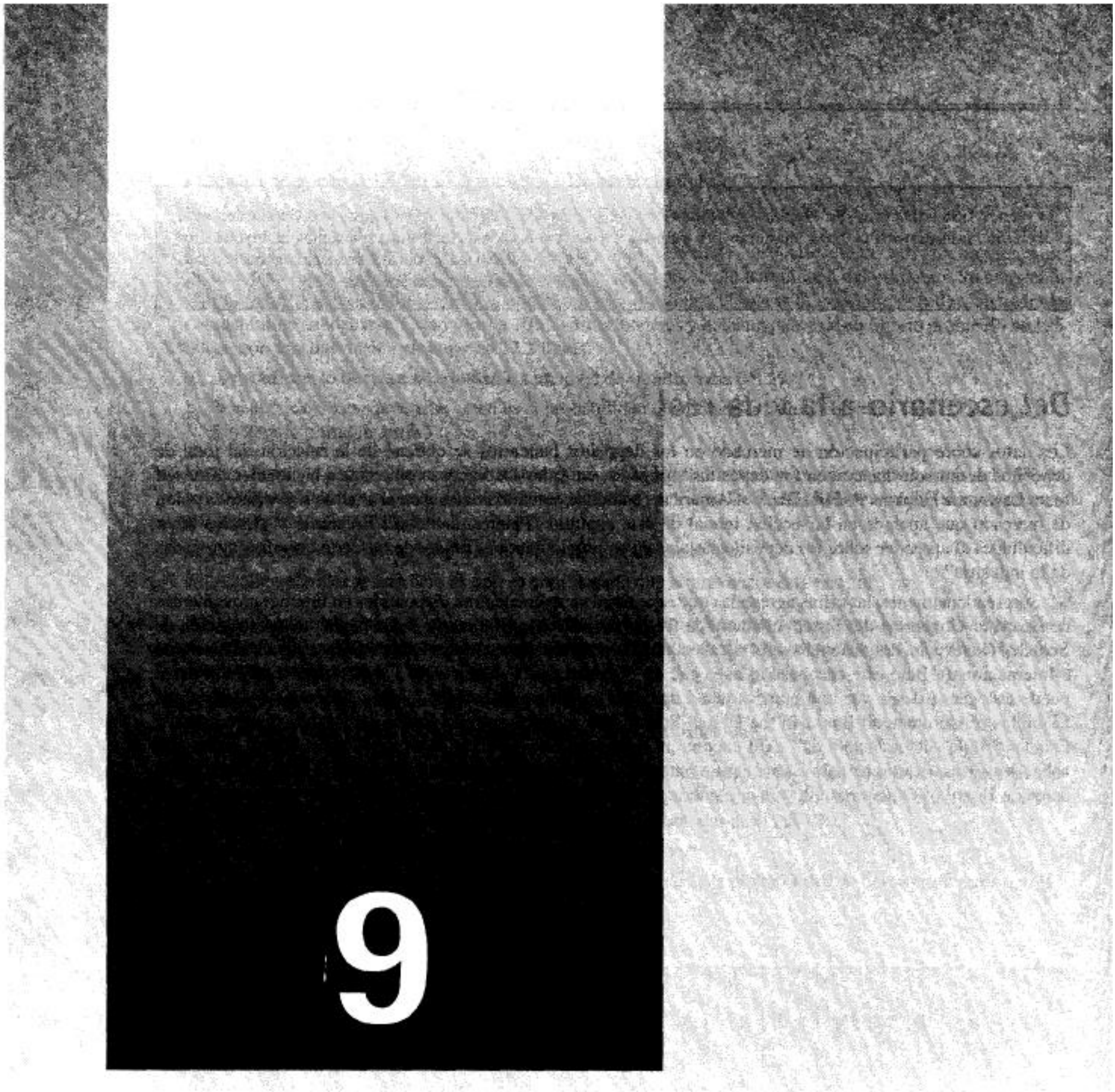
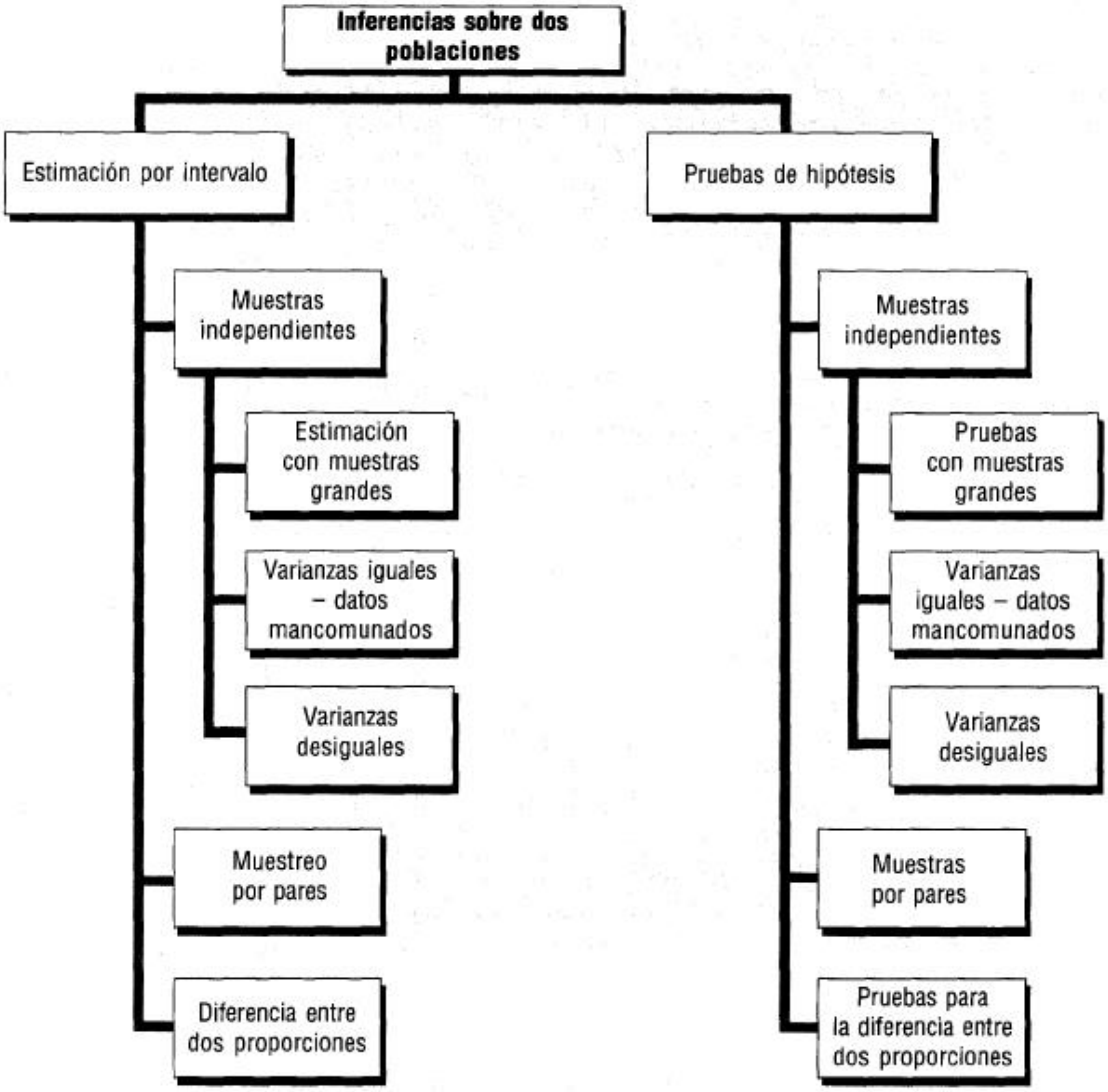


9

Pruebas cuando se tienen dos poblaciones

Plan del capítulo

Muchos problemas en los negocios requieren la comparación de dos poblaciones. Este capítulo analiza las situaciones en las cuales puede hacerse dicha comparación. Las ilustraciones muestran las circunstancias en las cuales es esencial comparar dos poblaciones y la forma apropiada como deben hacerse tales comparaciones.



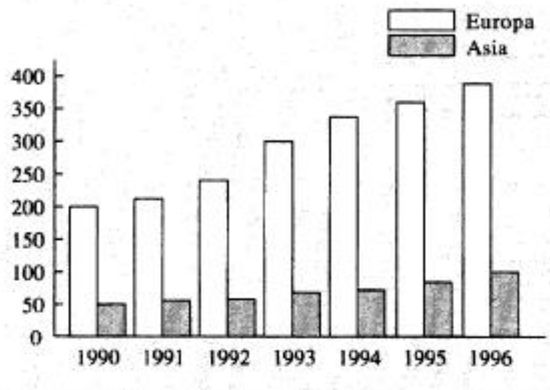
ESCENARIO

La revista *Fortune* (octubre de 1996) publicó una serie de artículos que discutían las tendencias del comercio exterior en los Estados Unidos. Estos artículos se centraron en las grandes sumas de dinero involucradas en las transacciones internacionales y en el papel que juegan Europa y Asia como lugares que compiten por la inversión de Estados Unidos. La inversión norteamericana en Europa fue de un total de US$364 billones en 1996, incrementando el 17% sobre el nivel máximo establecido justo al año anterior. Se presentó un incremento del 16% en la participación comercial de Estados Unidos en Asia, elevando el nivel de inversión en dicha región a más de US$100 billones. Estas tendencias se reflejan en la figura adjunta.

Los artículos también retaron la sabiduría convencional de que las empresas de Estados Unidos preferían invertir en las economías crecientes de Asia más que en el gran mercado existente en Europa. Sugirieron que los intereses comerciales nacionales todavía consideran que Europa ofrece una oportunidad más lucrativa para el crecimiento corporativo. Como analista internacional para su empresa, usted debe preparar un reporte comparativo detallado sobre las ventajas de la participación en cada una de estas dos áreas geopolíticas. Este reporte debe presentarse en una reunión ante muchos de los directivos de las divisiones de las empresas. El propósito de la reunión es decidir el curso futuro de la actividad inversionista en el extranjero para los próximos años. Esto requerirá que usted:

- Compare el rendimiento promedio sobre la inversión en Europa con la de Asia.
- Determine cuál área tiene un porcentaje menor de fallas en los proyectos de inversión.
- Estime el nivel de inversión promedio en Europa y Asia.
- Proporcione una comparación minuciosa de todas las medidas financieras relevantes en estos dos mercados extranjeros.

Inversiones de Estados Unidos en el exterior
(Billones de dólares)



9.1 Introducción

Los capítulos 7 y 8 mostraron cómo construir estimaciones por intervalo y probar hipótesis para una sola población. Este capítulo analiza cómo se pueden utilizar estas herramientas para comparar dos poblaciones. Por ejemplo, se puede querer construir una estimación por intervalo para la diferencia entre dos medias poblacionales o probar la hipótesis de que dos medias poblacionales son iguales. Muchas preguntas importantes pueden resolverse comparando dos poblaciones. ¿Cuál es la diferencia, si la hay, entre la durabilidad promedio de las botas de esquí elaboradas por North Slope y las producidas por Head? ¿Los trabajadores de una planta producen en promedio más que los trabajadores de una segunda planta? ¿Hay alguna diferencia entre la proporción de unidades defectuosas producidas por un método y las producidas por otro método alternativo?

El procedimiento exacto a seguir para la realización de estas pruebas depende de la técnica de muestreo que se utilice. Las muestras para pruebas con dos poblaciones pueden ser o 1) independientes o 2) por pares. Como el término lo indica, el muestreo independiente se realiza recolectando muestras independientes de cada población. Incluso las muestras no tienen que ser del mismo tamaño. Con el muestreo por pares, las observaciones de cada población tienen su correspondiente, lo cual se describirá más tarde detalladamente. Se comienza con una explicación sobre muestreo independiente.

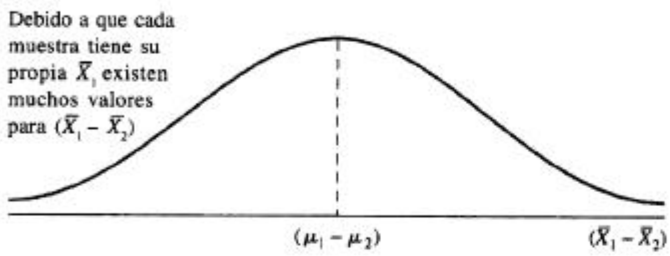
9.2 Estimación por intervalo en el caso de muestras independientes

Aquí el interés está en estimar la diferencia entre dos medias poblacionales ($\mu_1 - \mu_2$). El método apropiado depende de los tamaños de las muestras n_1 y n_2. Si tanto n_1 como n_2 son grandes (por lo menos 30), la técnica difiere en algo de aquella que se utiliza cuando alguno o ambos tamaños muestrales son de menos de 30. Veamos primero el método cuando ambas muestras son grandes.

A. Estimación con muestras grandes

La estimación puntual de la diferencia entre ($\mu_1 - \mu_2$) está dada por la diferencia entre las dos medias muestrales ($\bar{X}_1 - \bar{X}_2$). Ya que muchas muestras diferentes pueden tomarse de cada población, resulta toda una distribución de diferencias de estas medias muestrales. Si tanto n_1 y n_2 son grandes, la distribución de las diferencias entre las medias muestrales ($\bar{X}_1 - \bar{X}_2$) es una distribución normal centrada en ($\mu_1 - \mu_2$) tal y como lo muestra la figura 9.1

Figura 9.1
Distribución muestral de las diferencias de las medias muestrales ($\bar{X}_1 - \bar{X}_2$)



Dada esta distribución normal de las diferencias entre las medias muestrales, la desviación normal Z puede utilizarse para construir el intervalo. El procedimiento es muy parecido al desarrollado en el capítulo 7 para una sola población. Utilizando ($\bar{X}_1 - \bar{X}_2$) como la estimación puntual de la diferencia entre las dos medias poblacionales, se aplica un multiplicador de confianza para obtener los límites superior e inferior del intervalo.

Intervalo de confianza para la diferencia entre dos medias poblacionales - muestras grandes

$$\text{I.C. para } (\mu_1 - \mu_2) = (\bar{X}_1 - \bar{X}_2) \pm Z\sigma_{\bar{X}_1 - \bar{X}_2} \qquad [9.1]$$

en donde $\sigma_{\bar{X}_1 - \bar{X}_2}$ es el error estándar de las diferencias entre las medias muestrales. De la misma manera que con cualquier error estándar, $\sigma_{\bar{X}_1 - \bar{X}_2}$ mide la tendencia que tienen las diferencias entre las medias muestrales a variar. Se tiene que:

Error estándar de las diferencias entre medias muestrales	$\sigma_{\bar{x}_1-\bar{x}_2} = \sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}$	[9.2]

en donde σ_1^2 y σ_2^2 son las dos varianzas poblacionales. En el evento probable que σ_1^2 y σ_2^2 sean desconocidas, deben utilizarse las varianzas muestrales s_1^2 y s_2^2. El estimado del error estándar se vuelve entonces

Estimación del error estándar de la diferencia entre medias muestrales	$s_{\bar{x}_1-\bar{x}_2} = \sqrt{\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}}$	[9.3]

El intervalo para la diferencia en las medias muestrales es entonces

Intervalo de confianza cuando las varianzas poblacionales son desconocidas	$\text{I.C. para } (\mu_1 - \mu_2) = (\bar{X}_1 - \bar{X}_2) \pm Zs_{\bar{x}_1-\bar{x}_2}$	[9.4]

Vale la pena destacar que no se está interesado en el valor de cualquiera de las medias poblacionales, sino solamente en la diferencia que existe entre las dos medias poblacionales.

Transfer Trucking transporta remesas entre Chicago y Kansas City por dos rutas. Una muestra de 100 camiones enviados por la ruta del norte reveló un tiempo promedio de tránsito de $\bar{X}_N = 17.2$ horas con una desviación estándar de $S_N = 5.3$ horas, mientras que 75 camiones que utilizan la ruta del sur necesitaron un promedio de $\bar{X}_S = 19.4$ horas con una desviación estándar de $s_s = 4.5$ horas. Delmar, el despachador de Transfer Trucking, desea desarrollar un intervalo de confianza del 95% para la diferencia en el tiempo promedio entre estas dos rutas alternas.

Debido a que las desviaciones estándar de la población son desconocidas, el error estándar es

$$s_{\bar{x}_N-\bar{x}_S} = \sqrt{\frac{(5.3)^2}{100} + \frac{(4.5)^2}{75}} = 0.742$$

Un intervalo del 95% requiere un valor Z de 1.96. Entonces el intervalo es

$$\text{I.C. para } (\mu_N - \mu_S) = (17.2 - 19.4) \pm (1.96)(0.742)$$
$$-3.7 \leq (\mu_N - \mu_S) \leq -0.75$$

Los resultados pueden interpretarse como:

1. Delmar puede estar 95% seguro de que $(\mu_N - \mu_s)$ está entre –3.7 horas y –0.75 horas.

O, debido a que Delmar restó la media de la ruta del sur de la media de la ruta del norte y obtuvo números negativos,

2. Delmar puede tener un 95% de confianza en que la ruta del sur se toma entre 0.75 horas y 3.7 horas más.

Ejemplo 9.1

Charles Schwab, del servicio de corretaje de descuentos, recientemente instituyó dos programas de capacitación para los representantes de mercadeo telefónico contratados recientemente. Para probar la efectividad relativa de cada programa, a 45 representantes entrenados en el primer programa se les hizo la prueba de competencia. El puntaje promedio fue $\bar{X}_1 = 76.00$ con $s_1 = 13.50$ puntos. Las 40 personas entrenadas bajo el segundo programa reportaron un puntaje promedio de $\bar{X}_2 = 77.97$ y $s_2 = 9.05$. La gerencia desea saber si un programa de entrenamiento es más efectivo que el otro. Siendo usted la persona seleccionada para tomar esta determinación, usted decide construir un intervalo de confianza del 99% para hallar la diferencia entre los puntajes de competencia promedio de los empleados entrenados que se encuentran en cada programa. Usted también es responsable de recomendar cuál programa de entrenamiento deberá utilizar la compañía exclusivamente.

Solución

El error estándar de la diferencia entre la media muestral es

$$s_{\bar{x}_1-\bar{x}_2} = \sqrt{\frac{(13.50)^2}{45} + \frac{(9.05)^2}{40}} = 2.47$$

El intervalo del 99% es

$$\text{I.C. para } (\mu_1 - \mu_2) = (76.00 - 77.97) \pm (2.58)(2.47)$$
$$-8.34 \le (\mu_1 - \mu_2) \le 4.40$$

Interpretación

Usted puede tener un 99% de seguridad en que $(\mu_1 - \mu_2)$ está entre –8.34 puntos y 4.40 puntos. Dicho de otra manera, usted puede reportar con un 99% de confianza que los puntajes promedio de quienes recibieron entrenamiento con el programa 2 están entre 4.40 puntos menos y 8.34 puntos más que los entrenados con el programa 1. Es importante saber que debido a que el intervalo contiene cero, usted puede estar 99% seguro de que no hay diferencia en los puntajes promedio. Se puede concluir que cualquiera de los programas puede utilizarse para entrenar a los representantes recientemente contratados.

Ejercicios de la sección

1. Clark Insurance vende pólizas de seguro a los residentes en toda el área de Chicago. El propietario desea estimar la diferencia en las reclamaciones promedio de las personas que viven en zonas urbanas y las de aquellas que residen en los suburbios. De las 180 pólizas urbanas seleccionadas para la muestra, se reportó una reclamación promedio de US$2,025, con una desviación estándar de US$918. Las 200 pólizas urbanas revelaron una reclamación promedio de US$1,802 y s = US$512. ¿Qué dice al propietario un intervalo del 95% sobre las reclamaciones promedio registradas por estos dos grupos?

2. Dos procesos de producción se utilizan para producir tubos de acero. Una muestra de 100 tubos tomada del primer proceso de producción tiene una longitud promedio de 27.3 pulgadas y $s = 10.3$ pulgadas. Las cifras correspondientes para los 100 tubos producidos por el segundo método son 30.1 y 5.2. ¿Qué revela un intervalo del 99% sobre la diferencia en las longitudes promedio de los tubos producidos por estos dos métodos?

3. Chapman Industries utiliza dos dispositivos de contestador telefónico. Chuck Chapman desea determinar si los clientes que llaman se dejan esperando, en promedio, más en un sistema que en el otro. Si 75 llamadas en el primer sistema reportaron un tiempo de espera promedio de 25.2 segundos con $s = 4.8$ segundos y 70 llamadas del segundo sistema dieron una media de 21.3 segundos con s= 3.8 segundos, ¿qué recomendación proporcionaría usted a Chuck si él desea minimizar el tiempo de espera con base en un intervalo del 90% estimando de la diferencia de tiempo de espera promedio?

4. Se utilizan dos diseños de producción para fabricar cierto producto. El tiempo promedio requerido para producir el producto utilizando el viejo diseño fue de 3.51 días con $s = 0.79$ días. El nuevo diseño requirió un promedio de 3.32 días con $s = 0.73$ días. Muestras de igual tamaño de 150 se utilizaron en ambos diseños. ¿Qué revela un intervalo del 99% sobre la diferencia entre los tiempos promedio que se requieren para hacer el producto? Determine cuál diseño debería utilizarse.
5. Explique exactamente lo que en realidad mide el error estándar de la diferencia entre medias muestrales.

B. Estimación con muestras pequeñas: La distribución *t*

Si cualquier muestra es pequeña (menor que 30), no se puede asumir que la distribución de las diferencias en las medias muestrales $(\bar{X}_1 - \bar{X}_2)$ se ajusta a una distribución normal. Por tanto, no puede utilizarse la desviación normal Z. Si

1. Las poblaciones están distribuidas normalmente o distribuidas casi normalmente, y
2. Las varianzas poblacionales son desconocidas,

Se debe regresar a la distribución t.

Para aplicar adecuadamente la distribución t, también se debe determinar si las varianzas de las dos poblaciones son iguales. ¿Cómo se puede asumir que tales varianzas son iguales si, como se anotó anteriormente, no se sabe cuáles son? Muchos procesos de las líneas de ensamble utilizan máquinas para llenar los contenedores de producto tales como latas, botellas y cajas. Cuando se ajustan periódicamente las máquinas para garantizar la operación adecuada, se asume que el nivel de contenido promedio ha variado pero no se conoce la varianza en los niveles de contenido antes y después del ajuste; aunque sean desconocidas, siguen sin cambio alguno. Otros ejemplos también prevalecen, en los cuales la suposición de varianzas iguales es razonable. Más tarde, en este capítulo, se presentará un método para probar la igualdad de las dos varianzas. Por ahora, se analizarán los intervalos de confianza cuando 1) se asume que las varianzas son iguales y cuando 2) no se asume que las varianzas sean iguales.

1. Suposición de que las varianzas $\sigma_1^2 = \sigma_2^2$ son iguales pero desconocidas Si las varianzas de las dos poblaciones son iguales, existe alguna varianza σ^2 común a ambas poblaciones. Es decir $\sigma_1^2 = \sigma_2^2 = \sigma^2$. Sin embargo, debido al error de muestreo, si se toma una muestra de cada población, las dos varianzas de la muestra probablemente diferirán una de la otra así como de la varianza común σ^2. Pero debido a que las poblaciones tienen una varianza común, los datos de ambas muestras pueden mancomunarse (*pooled*) para obtener un solo estimado de σ^2. Esto se hace calculando el promedio ponderado de las dos varianzas de la muestra, en donde los pesos son los grados de libertad $n - 1$ para cada muestra. Esta estimación **mancomunada (pooled estimate)** de la varianza poblacional común s_p^2 es:

Estimado mancomunado de la varianza común a ambas poblaciones

$$s_p^2 = \frac{s_1^2(n_1 - 1) + s_2^2(n_2 - 1)}{n_1 + n_2 - 2} \qquad [9.5]$$

El intervalo de confianza para la diferencia entre las dos medias poblacionales se halla entonces con una distribución t con $n_1 + n_2 - 2$ grados de libertad.

Intervalo para la diferencia entre medias poblacionales cuando $\sigma_1^2 = \sigma_2^2$ desconocidas

$$\text{I.C. para } (\mu_1 - \mu_2) = (\bar{X}_1 - \bar{X}_2) \pm t\sqrt{\frac{s_p^2}{n_1} + \frac{s_p^2}{n_2}} \qquad [9.6]$$

En la cafetería de los estudiantes, una máquina expendedora de bebidas dispensa bebidas en tazas de papel. Una muestra de 15 tazas da una media de 15.3 onzas con una varianza de 3.5. Después de ajustar la máquina, una muestra de 10 tazas produce un promedio de 17.1 onzas con una varianza de 3.9. Si se asume que s^2 es constante antes y después del ajuste, construya un intervalo de confianza del 95% para la diferencia entre los contenidos promedio de llenado. Se asume que las cantidades dispensadas están distribuidas normalmente.

Entonces,

$$s_p^2 = \frac{3.5(14) + 3.9(9)}{15 + 10 - 2}$$
$$= 3.66$$

Con $\alpha = 0.05$ (un nivel de confianza del 95%) y $n_1 + n_2 - 2 = 23$ g.l., la tabla t indica un valor de 2.069.

$$\text{I.C. para } \mu_1 - \mu_2 = (15.3 - 17.1) \pm 2.069\sqrt{\frac{3.66}{15} + \frac{3.66}{10}}$$
$$= -1.8 \pm 1.61$$
$$-3.41 \leq \mu_1 - \mu_2 \leq -0.19$$

Restando el nivel de contenido promedio después del ajuste (17.1), del nivel de contenido promedio antes del ajuste (15.3), resultan valores negativos tanto para los extremos inferiores como superiores del intervalo. Es decir, el intervalo no contiene cero. Por consiguiente se puede tener un nivel de confianza del 95% en el que el ajuste incrementó el nivel de contenido promedio entre 0.19 onzas y 3.41 onzas.

Ejemplo 9.2

Las negociaciones salariales entre su empresa y el sindicato de sus trabajadores están a punto de romperse. Existe un desacuerdo considerable sobre el nivel salarial promedio de los trabajadores en la planta de Atlanta y en la planta de Newport News, Virginia. Los salarios fueron fijados por el antiguo acuerdo laboral de hace tres años y se basan estrictamente en la antigüedad. Debido a que los salarios están controlados muy de cerca por el contrato laboral, se asume que la variación en los salarios es la misma en ambas plantas y que los salarios están distribuidos normalmente. Sin embargo, se siente que existe una diferencia entre los niveles salariales promedio debido a los patrones de antigüedad diferentes entre las dos plantas.

El negociador laboral que representa a la gerencia desea que usted desarrolle un intervalo de confianza del 98% para estimar la diferencia entre los niveles salariales promedio. Si existe una diferencia en las medias, deben hacerse ajustes para hacer que los salarios más bajos alcancen el nivel de los más altos. Dados los siguientes datos, ¿qué ajustes se requieren, si es el caso?

Solución

Las muestras de trabajadores tomadas de cada planta revelan la siguiente información.

Planta de Atlanta	Planta de Newport News
$n_A = 23$	$n_N = 19$
$\bar{X}_A$ = US\$17.53 por hora	$\bar{X}_N$ = US\$15.50
$s_A^2 = 92.10$	$s_A^2 = 87.10$

Entonces:

$$s_p^2 = \frac{92.10(22) + 87.1(18)}{23 + 19 - 2}$$
$$= 89.85$$

Dado que $\alpha = 0.02$ y g.l. = 23 + 19 – 2 = 40, la tabla F en el apéndice III revela un valor t de 2.423.

$$\text{I.C. para } \mu_1 - \mu_2 = (17.53 - 15.5) \pm 2.423\sqrt{\frac{89.85}{23} + \frac{89.85}{19}}$$
$$= 2.03 \pm 7.12$$
$$-5.09 \leq \mu_1 - \mu_2 \leq 9.15$$

Interpretación

Se puede estar 98% seguro de que el salario promedio en Atlanta está entre US\$9.15 más que en Newport News y US\$5.09 menos que en Newport News. Debido a que este intervalo contiene el valor US\$0, la conclusión es que es posible que no exista diferencia. No se garantiza ajuste.

2. Varianzas desiguales Si las varianzas de la población son desiguales o no existe evidencia para asumir una igualdad, los procedimientos descritos anteriormente no se aplican directamente porque la distribución de las diferencias entre las medias muestrales no se ajustan a una distribución t con $n_1 + n_2 - 2$ g.l. De hecho, no se ha encontrado que una distribución exacta describa este proceso de muestreo, y sólo se han desarrollado aproximaciones. Se ha propuesto una aproximación tal que utilice el estadístico t con g.l. levemente alterados. En el evento que $\sigma_1^2 \neq \sigma_2^2$, se encuentra que los grados de libertad son

Grados de libertad cuando las varianzas poblacionales no son iguales

$$\text{g.l.} = \frac{(s_1^2/n_1 + s_2^2/n_2)^2}{(s_1^2/n_1)^2/(n_1 - 1) + (s_2^2/n_2)^2/(n_2 - 1)} \qquad [9.7]$$

Debido a que g.l. se calcula de esta manera alterada, el estadístico t se simboliza con t'. El intervalo de confianza se ha calculado entonces como

Intervalo para la diferencia entre medias poblacionales

$$\text{I.C. para } \mu_1 - \mu_2 = (\bar{X}_1 - \bar{X}_2) \pm t'\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} \qquad [9.8]$$

The Wall Street Journal describió dos programas de entrenamiento utilizados por IBM. Doce ejecutivos a quienes se les dio el primer tipo de entrenamiento obtuvieron un promedio de 73.5 en la prueba de competencia. Aunque el artículo de noticias no reportó la desviación estándar para estos 12 empleados, se asume que la varianza en los puntajes para este grupo fue de 100.2. Quince ejecutivos a quienes se les administró el segundo programa de entrenamiento obtuvieron un promedio de 79.8. Se asume una varianza de 121.3 para este segundo grupo. Haga un intervalo de confianza del 95% para la diferencia en los puntajes promedio para todos los ejecutivos que ingresaron a estos programas:

$$\text{g.l.} = \frac{(100.2/12 + 121.3/15)^2}{(100.2/12)^2/11 + (121.3/15)^2/14}$$
$$= 24.55$$

Si g.l. es fraccionario, se aproxima hacia abajo, hacia el entero inmediatamente anterior. Así, g.l. = 24. Un intervalo de confianza del 95% con 24 grados de libertad requiere un valor t' de 2. 064.

$$
\begin{aligned}
\text{I.C. para } \mu_1 - \mu_2 &= (\bar{X}_1 - \bar{X}_2) \pm t'\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} \\
&= (73.5 - 79.8) \pm 2.064\sqrt{\frac{100.2}{12} + \frac{121.3}{15}} \\
&= -6.3 \pm 8.36 \\
-14.66 \leq \mu_1 &- \mu_2 \leq 2.06
\end{aligned}
$$

Debido a que el intervalo contiene cero, no existe una fuerte evidencia de que exista diferencia alguna en la efectividad de los programas de entrenamiento.

Ejemplo 9.3

Acme Ltd. vende dos tipos de amortiguadores de caucho para coches de bebés. Las pruebas de desgaste para medir la durabilidad revelaron que 13 amortiguadores del tipo 1 duraron un promedio de 11.3 semanas, con una desviación estándar de 3.5 semanas; mientras que 10 del tipo 2 duraron un promedio de 7.5 semanas, con una desviación estándar de 2.7 semanas. El tipo 1 es más costoso para fabricar y el CEO (Director Ejecutivo) de Acme no desea utilizarlo a menos que tenga un promedio de duración de por lo menos ocho semanas más que el tipo 2. El CEO tolerará una probabilidad de error de sólo el 2%. No existe evidencia que sugiera que las varianzas de la duración de los dos productos sean iguales.

Solución

$$
\begin{aligned}
\bar{X}_1 &= 11.3 \qquad & \bar{X}_2 &= 7.5 \\
n_1 &= 13 & n_2 &= 10 \\
s_1 &= 3.5 & s_2 &= 2.7
\end{aligned}
$$

$$
\begin{aligned}
\text{g.l.} &= \frac{(s_1^2/n_1 + s_2^2/n_2)^2}{(s_1^2/n_1)^2/(n_1 - 1) + (s_2^2/n_2)^2/(n_2 - 1)} \\
&= \frac{[(3.5)^2/13 + (2.7)^2/10]^2}{\dfrac{[(3.5)^2/13]^2}{12} + \dfrac{[(2.7)^2/10]^2}{9}} \\
&= 20.99 \approx 20
\end{aligned}
$$

Un intervalo de confianza del 98% (α=0.02) con 20 g.l. requiere un valor t de 2.528.

$$
\begin{aligned}
\text{I.C. para } \mu_1 - \mu_2 &= (11.3 - 7.5) \pm 2.528\sqrt{\frac{(3.5)^2}{13} + \frac{(2.7)^2}{10}} \\
&= 3.8 \pm 3.3 \\
0.5 \leq \mu_1 &- \mu_2 \leq 7.1 \text{ semanas}
\end{aligned}
$$

Interpretación

Acme puede estar 98% seguro de que el tipo 1 dura entre 0.5 y 7.1 semanas más que el tipo 2. Debido a que la diferencia requerida de ocho semanas no está en el intervalo, el CEO puede estar 98% seguro de que no desea utilizar el tipo 1.

Ejercicios de la sección

6. ¿Qué condiciones debe reunir la distribución t antes que pueda utilizarse?
7. Diecisiete latas de Croc Aid presentan una media de 17.2 onzas, con una desviación estándar de 3.2 onzas, y 13 latas de Energy Pro producen una media de 18.1 onzas y $s = 2.7$ onzas. Asumiendo varianzas iguales y distribuciones normales en los pesos de la población, ¿qué conclusiones se pueden sacar respecto a la diferencia en los pesos promedio, con base en un intervalo de confianza del 98%?
8. Grow-rite vende un fertilizante comercial producido en dos plantas de Atlanta y Dallas. Las quejas recientes de un cliente sugieren que a los envíos de Atlanta les falta peso, comparados con los envíos de Dallas. ¿Si 10 cajas de la planta de Atlanta tienen un promedio de 96.3 libras con $s = 12.5$ y 15 cajas de la planta de Dallas tienen un promedio de 101.1 con $s = 10.3$, un intervalo de confianza del 99% confirma esta queja? Se asumen varianzas iguales.
9. Opus, Inc. ha desarrollado un proceso para producir oro a partir del agua de mar. Quince galones tomados del océano Pacífico produjeron una media de 12.7 onzas de oro por galón con $s = 4.2$ onzas, y 12 galones del océano Atlántico produjeron cifras similares de 15.9 y 1.7. Con base en un intervalo del 95%, ¿cuál es su estimado de la diferencia en las onzas promedio de oro provenientes de estas dos fuentes? No existe razón para asumir que las varianzas son iguales.
10. Ralphie comienza la universidad en otoño próximo. Toma muestras de apartamentos en los extremos norte y sur de la ciudad para ver si hay alguna diferencia en las rentas promedio. Los apartamentos que están en el norte reportan rentas de US\$600, US\$650, US\$530, US\$800, US\$750, US\$700 y US\$750, y los del sur reportaron rentas de US\$500, US\$450, US\$800, US\$650, US\$500, US\$500, US\$450, y US\$400. Si no existe evidencia alguna de que las varianzas sean iguales, ¿qué dice a Ralphie un intervalo del 99% sobre la diferencia en las rentas promedio?
11. Bigelow Products desea desarrollar un intervalo del 95% para estimar la diferencia en las ventas semanales promedio en dos mercados objetivo. Una muestra de 9 semanas en el mercado 1 produjo una media y una desviación estándar en cientos de dólares de 5.72 y 1.008 respectivamente. Las cifras comparables para el mercado 2, con base en una muestra de 10 semanas, fueron de 8.72 y 1.208. ¿Si se asumen varianzas iguales, qué resultados reportan?
12. U.S. Manufacturing compra materias primas de dos proveedores. La gerencia está preocupada sobre los retrasos en la producción debido a que no se reciben los envíos a tiempo. Una muestra de 10 envíos del proveedor A tienen un tiempo de entrega promedio de 6.8 días y $s= 2.57$ días mientras que 12 envíos del proveedor B tienen un promedio de 4.08 días y $s = 1.93$. Si no se pueden asumir varianzas iguales, ¿qué recomendación haría con base en un intervalo del 90% para la diferencia en los tiempos promedio de entrega?

9.3 Estimación del intervalo con muestras pareadas

También denominadas pares correspondientes, las **muestras pareadas** involucran un procedimiento en el cual varios pares de observaciones se equiparan de la manera más próxima posible, en términos de características relevantes. Los dos grupos de observaciones son diferentes sólo en un aspecto o "tratamiento". Toda diferencia subsiguiente en los dos grupos se atribuye a dicho tratamiento. Las ventajas de las muestras pareadas son: 1) pueden utilizarse muestras más pequeñas, 2) se encuentran varianzas más pequeñas, 3) menos grados de libertad se pierden en el análisis, y 4) resulta un error de muestreo más pequeño (la variación entre observaciones se reduce debido a que corresponden de la forma más próxima posible). Por este motivo se prefieren las muestras pareadas si su aplicación es posible.

Muestras pareadas Se llaman pares correspondientes dos observaciones que son lo más similares posible entre sí. Sólo difieren en un aspecto relevante.

Supongamos que un investigador médico desea probar los efectos de nuevos medicamentos sobre los niveles de presión sanguínea en los pacientes. Veinte personas en un grupo se les hace corresponder de la manera más próxima posible con 20 personas de un segundo grupo en términos de peso, edad, sexo, nivel de actividad, colesterol y cualquier otro factor que pueda afectar la presión sanguínea. Tendremos entonces 20 pares de "gemelos idénticos". A uno de los grupos se les proporciona una nueva droga, y el otro recibe otra medicina. Sólo el investigador sabe cuál grupo recibe cada medicamento. Toda diferencia subsiguiente, buena o mala, en las presiones sanguíneas de los "gemelos" se atribuye a las medicinas debido a que se han "corregido" todos los otros factores relevantes que afectan la presión sanguínea (como la edad y el sexo); es decir, se han elaborado pares de observaciones (gemelos) que son idénticos en términos de los demás factores relevantes.

Otro método para utilizar muestras pareadas involucra el análisis de las mismas observaciones antes y después del tratamiento. Una práctica común en la industria es evaluar a los empleados en los programas de capacitación. Entonces se tiene un puntaje "antes" y uno "después" de la capacitación en cada observación (empleado). Todo cambio en el puntaje puede atribuirse a esta formación adicional. (Al investigador debería prevenírsele que algo de sesgo puede afectar el muestreo debido a que los empleados pueden ofenderse por sentirse evaluados o pueden recordar sus respuestas de la primera evaluación y tratar de ser consistentes en las mismas).

Las muestras pareadas tienen ciertas ventajas ya que las muestras más pequeñas con frecuencia conducirán a resultados más precisos; si controla los demás factores importantes, el investigador no tendrá que tomar muestras grandes para reducir el error de muestreo.

Para ilustrar se asume que se tienen puntajes de la prueba de 10 empleados antes y después de habérseles impartido capacitación laboral adicional. Los puntajes aparecen en la tabla 9.1.

Tabla 9.1
Puntajes de la prueba antes y después de la capacitación en el trabajo

Empleado	Puntaje antes de la capacitación en el trabajo	Puntaje después de la capacitación en el trabajo	d_i	d_i^2
1	9.0	9.2	-0.2	0.04
2	7.3	8.2	-0.9	0.81
3	6.7	8.5	-1.8	3.24
4	5.3	4.9	0.4	0.16
5	8.7	8.9	-0.2	0..04
6	6.3	5.8	0.5	0.25
7	7.9	8.2	-0.3	0.09
8	7.3	7.8	-0.5	0.25
9	8.0	9.5	-1.5	2.25
10	7.5	8.0	-0.5	0.25
	74.0	79.0	-5.0	7.38

Sea d_i la diferencia entre todo par correspondiente. La media de las diferencias entre todos los pares sería entonces:

Diferencia promedio entre las observaciones pareadas

$$\bar{d} = \frac{\Sigma d_i}{n} \qquad [9.9]$$

$$= \frac{-5.0}{10} = -0.5$$

La desviación estándar de estas diferencias es:

Desviación estándar de las diferencias entre las observaciones pareadas

$$s_d = \sqrt{\frac{\Sigma d_i^2 - n\bar{d}^2}{n-1}} \qquad [9.10]$$

$$= \sqrt{\frac{7.38 - 10(-0.5)^2}{9}}$$

$$= 0.736$$

Como $n < 30$ y la desviación estándar de las diferencias de los puntajes σ_d es desconocida, se requiere el uso del estadístico t. Si n hubiera sido mayor que 30 o σ_d fuera conocido, el estadístico Z se hubiera podido utilizar. Además, debe asumirse que los valores d siguen una distribución normal. La distribución de los puntajes puros en sí misma es inmaterial, pero los valores d deben ser normales.

Entonces para un nivel de confianza del 90% y $n - 1 = 9$ grados de libertad, un intervalo de confianza para la media de la diferencia en los puntajes antes y después de la capacitación es:

Intervalo para la diferencia entre medias; observaciones pareadas

$$\text{I.C. para } \mu_d = \bar{d} \pm t\frac{s_d}{\sqrt{n}} \qquad [9.11]$$

$$= -0.5 \pm (1.833)\frac{0.736}{\sqrt{10}}$$

$$-0.927 \leq \mu_d \leq -0.073$$

Debido a que se restaron los puntajes posteriores al entrenamiento de los puntajes anteriores al entrenamiento, produciendo valores negativos, se puede estar 90% seguro de que la media de los puntajes posteriores al entrenamiento está entre 0.073 puntos y 0,927 puntos más alto.

Ejemplo 9.4

Vicki Peplow, directora regional de pagos de asistencia médica para Aetna Insurance, en Peoria, Illinois, constató que dos hospitales diferentes parecían cobrar cantidades ampliamente diferentes por el mismo procedimiento médico. Ella recolectó observaciones sobre costos de facturación para 15 procedimientos idénticos en cada hospital, y construyó un intervalo de confianza del 95% para la diferencia entre los costos promedio presentados por cada hospital. Se utilizaron muestras pareadas porque Vicki corrigió todos los demás factores relevantes distintos al costo.

Si existe una diferencia, La señora Peplow planea reportar este asunto a las autoridades de asistencia médica Medicare. ¿Deberá ella presentar el informe?

Hospital 1	Hospital 2	d_i	d_i^2
465	512	-47	2,209
532	654	-122	14,884
426	453	-27	729
543	521	22	484
587	632	-45	2,025
537	418	119	14,161
598	587	11	121
698	376	322	103,684
378	529	-151	22,801
376	517	-141	19,881

(Continúa)

Hospital 1	Hospital 2	d_i	d_i^2
524	476	48	2,304
387	519	-132	17,424
429	587	-158	24,964
398	639	-241	58,081
412	754	-342	116,964
		$-884 = \Sigma d_i$	$400{,}716 = \Sigma d_i^2$

Solución

Dados los datos anteriores se tiene que:

$$\bar{d} = \frac{\Sigma d_i}{n} = \frac{-884}{15} = -58.93$$

$$s_d = \sqrt{\frac{\Sigma d_i^2 - n\bar{d}^2}{n-1}} = \sqrt{\frac{400{,}716 - (15)(-58.93)^2}{15-1}}$$
$$= 157.8$$

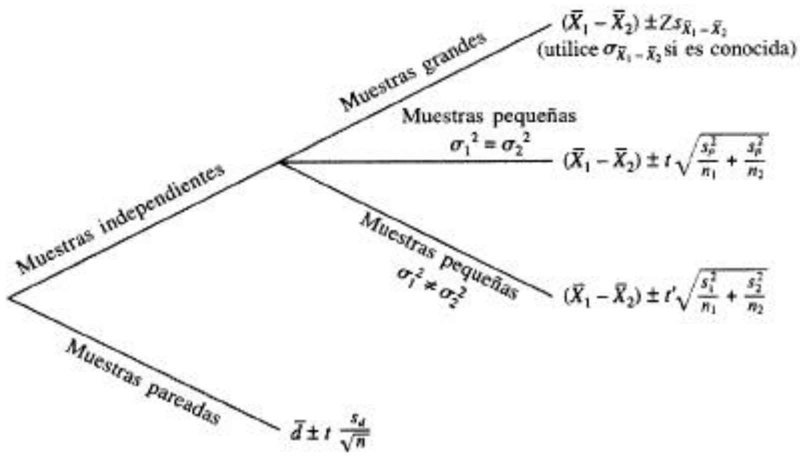
$$\text{I.C. para } \mu_1 - \mu_2 = \bar{d} \pm (t)s_d/\sqrt{n}$$
$$= -58.93 \pm (2.145)157.8/\sqrt{15}$$
$$= -58.93 \pm 87.4$$
$$-146.33 \le \mu_1 - \mu_2 \le -28.47$$

Interpretación

Debido a que resultaron valores con signos contrarios cuando se restaron los cargos del segundo hospital de los del primer hospital, Vicki puede estar 95% segura de que los hospitales están cobrando aproximadamente lo mismo por los mismos servicios. Un reporte para los administradores de la asistencia médica apropiados no parece adecuado.

La figura 9.2 ayudará a decidir qué fórmula y qué método deberían utilizarse en la construcción de estos intervalos.

Figura 9.2
Cálculo de los intervalos para la diferencia entre dos medias poblacionales



Ejercicios de la sección

13. Rankin Associates recibirá licitaciones de dos empresas constructoras para un trabajo de remodelación en su oficina principal. La decisión sobre cuál oferta aceptarán dependerá en parte de los tiempos promedio de terminación de contratos similares de cada compañía. Se han recolectado y agrupado por pares los datos para varios trabajos de remodelación. Con base en un nivel de confianza del 99%, ¿cuál compañía recomendaría usted?

Compañía 1	Compañía 2	Compañía 1	Compañía 2
10.0	9.2	12.5	7.2
12.2	10.0	7.3	8.4
15.3	9.2	9.4	10.5
9.6	10.5	8.7	6.2
8.6	9.5	9.1	8.1
9.4	8.4		

14. En su último viaje a Las Vegas, Lucky Louie jugó 15 juegos de blackjack en el casino del hotel Golden Nugget, y 15 juegos en el del hotel Flamingo. Para comparar sus ganancias promedio, Louie restó su "producto" del Flamingo del "producto" de el Nugget y encontró que $\Sigma d_i = 40$ y $\Sigma d_i^2 = 415$. Utilizando un intervalo de confianza del 95%, ¿en qué lugar debería jugar con más frecuencia Louie si desea maximizar sus ganancias promedio?

15. La maratón anual se realizará este fin de semana. Usted ha registrado su tiempo durante el transcurso de las últimas cinco carreras utilizando dos tipos de tenis para correr. Dispuesto a ganar la carrera este año, usted desea estimar la diferencia en sus tiempos promedio utilizando estos dos tipos de zapatos deportivos. Al restar los tiempos que usted registró utilizando SpeedBurst Shoes de los tiempos que registró utilizando los zapatos RocketMaster obtuvo en minutos: $\Sigma d_i = 8$ y $\Sigma d_i^2 = 24$. A un nivel de confianza del 90%, ¿cuál zapato prefiere?

16. Como analista de inversión usted debe comparar los rendimientos promedio de dos tipos de títulos que su cliente está interesado en comprar. Los siguientes datos son sobre 12 títulos de cada tipo. ¿Qué recomendación haría con base en un intervalo de confianza del 99% sobre la diferencia entre los rendimientos promedio de cada tipo de título? Se asume que las observaciones son pareadas.

Título 1	Título 2	Título 1	Título 2
3.21%	6.39%	6.58%	4.58%
6.50	8.69	4.58	4.00
8.25	7.89	7.80	7.80
9.32	9.58	4.60	9.88
5.26	6.57	5.89	6.58
4.58	7.48	6.66	7.89

17. Los salarios mensuales iniciales, en miles de dólares, para 12 graduados en administración de la Tech U, se comparan con los de la State U, utilizando los datos que aparecen a continuación. Desarrolle e interprete un intervalo del 95% para hallar la diferencia en los salarios promedio iniciales obtenidos al restar los salarios de State U de los de Tech U.

Tech	State	Tech	State
3.7	5.6	2.5	8.8
3.6	6.8	3.5	9.5
5.2	8.5	3.9	7.5
1.2	6.5	8.2	6.5
1.6	5.5	4.5	4.5
5.2	4.8	1.2	8.7

18. Utilizando los datos del problema anterior, calcule e interprete el intervalo del 95% para hallar la diferencia en los salarios promedio obtenidos al restar los salarios de Tech de los de State. ¿Qué diferencias encuentra?

9.4 Intervalos de confianza para la diferencia entre dos proporciones

A veces surgen situaciones en las cuales es necesario comparar las proporciones de dos poblaciones diferentes. Las empresas continuamente analizan la proporción de productos defectuosos producidos por métodos diferentes. Investigaciones médicas se interesan en la proporción de hombres que sufren de ataques al corazón en oposición al porcentaje de mujeres. En general, muchos asuntos comerciales están determinados mediante la estimación de proporciones relativas de dos poblaciones.

El procedimiento ya debería ser familiar para usted. El error estándar de la diferencia entre dos proporciones muestrales ($p_1 - p_2$) se estima mediante la fórmula:

Error estándar de la diferencia entre dos proporciones muestrales

$$s_{p_1-p_2} = \sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}} \qquad [9.12]$$

El término $s_{p_1-p_2}$ reconoce que si se tomaran varios pares de muestras de cada población ($p_1 - p_2$) variaría. La fórmula (9.12) tiene en cuenta dicha variación. El intervalo de confianza es:

Intervalo para la diferencia entre proporciones poblacionales

$$\text{I.C. para } \pi_1 - \pi_2 = (p_1 - p_2) \pm (Z)s_{p_1-p_2} \qquad [9.13]$$

Una empresa realiza un estudio para determinar si el ausentismo de los trabajadores en el turno del día es diferente al de los trabajadores del turno de la noche. Se realiza una comparación de 150 trabajadores de cada turno. Los resultados muestran que 37 trabajadores diurnos han estado ausentes por lo menos cinco veces durante el año anterior, mientras que 52 trabajadores nocturnos han faltado por lo menos cinco veces. ¿Qué revelan estos datos sobre la tendencia al ausentismo entre los trabajadores? Calcule un intervalo de confianza del 90% para la diferencia entre las proporciones de trabajadores de los dos turnos que faltaron cinco veces o más.

$$p_1 = \frac{37}{150} = 0.25, \qquad p_2 = \frac{52}{150} = 0.35$$

$$\begin{aligned}
s_{p_1-p_2} &= \sqrt{\frac{(0.25)(0.75)}{150} + \frac{(0.35)(0.65)}{150}} \\
&= 0.0526 \\
\text{I.C. para } \pi_1 - \pi_2 &= (p_1 - p_2) \pm (Z)s_{p_1-p_2} \\
&= (0.25 - 0.35) \pm (1.65)(0.0526) \\
&= -0.10 \pm 0.087 \\
-18.7\% \leq (\pi_1 - \pi_2) &\leq -1.3\%
\end{aligned}$$

Debido a que la proporción de trabajadores nocturnos que se ausentaron cinco veces o más (p_2) se restó de la proporción de trabajadores diurnos que se ausentaron, la empresa puede estar 90% segura de que la proporción de trabajadores nocturnos ausentes en cinco o más oportunidades está entre 1.3% y 18.7% más alta que los del turno diurno.

El ejemplo 9.5 proporciona otro panorama de una estimación por intervalo para la diferencia entre proporciones.

Ejemplo 9.5

Su empresa utiliza dos máquinas diferentes para cortar los disfraces Spandex que utilizan los actores de la obra musical Ice Capades. Se han presentado problemas en cuanto al ajuste apropiado, debido al funcionamiento de las máquinas. Como director de control de calidad su trabajo es estimar la diferencia en la proporción de defectos producidos por cada máquina. Se tomaron muestras de tamaños $n_1 = 120$ y $n_2 = 95$: la primera máquina produjo 38% de defectos y la segunda, 43% de defectos. Fijar α en el 5%. Si la evidencia sugiere que la diferencia en la proporción de defectos excede del 5%, todos los disfraces se producirán en la máquina que parezca tener una tasa de defectos menor. ¿Qué decisión tomará usted?

Solución

$$
\begin{aligned}
s_{p_1-p_2} &= \sqrt{\frac{(0.38)(0.62)}{120} + \frac{(0.43)(0.57)}{95}} \\
&= 0.0674
\end{aligned}
$$

$$
\begin{aligned}
\text{I.C. para } \pi_1 - \pi_2 &= (0.38 - 0.43) \pm (1.96)(0.0674) \\
&= -0.05 \pm 0.1321 \\
&= -0.1821 \leq (\pi_1 - \pi_2) \leq 0.0821
\end{aligned}
$$

Interpretación

Usted puede estar 95% seguro de que la proporción de defectos producidos por la primera máquina está entre 18.21% menos y 8.21% más que la segunda máquina. Debido a que el intervalo contiene cero, no hay evidencia de que exista alguna diferencia en la proporción de defectos producidos por cada una de las máquinas. Su decisión es que como no hay diferencia en cuanto a cuál máquina utilizar, usted puede continuar utilizando ambas.

Ejercicios de la sección

19. De 150 hombres y 130 mujeres, 27% y 35% respectivamente afirmaron que utilizaban tarjetas de crédito para comprar regalos de navidad. Calcule e interprete el intervalo de confianza del 99% para la diferencia entre la proporción de hombres y mujeres que optaron por el crédito.

20. Los registros muestran que de 1000 estudiantes de fuera del estado, 40% fueron a casa para las vacaciones de primavera al tiempo que el 47% de los 900 estudiantes del estado también lo hicieron. ¿Cuál es el intervalo de confianza del 95% para la diferencia en la proporción de estos estudiantes que fueron a casa?

21. De los 50 estudiantes graduados 10 fueron a un clima cálido para las vacaciones de primavera, mientras que 24 de los 75 no graduados hicieron lo mismo. Construya el intervalo del 95% de confianza.

9.5 Selección del tamaño apropiado de la muestra

Con frecuencia es necesario determinar el tamaño apropiado de una muestra para realizar determinado estudio. Al igual que con una población aislada, el tamaño de la muestra requerido depende de: 1) las varianzas poblacionales, y 2) el grado de exactitud deseado. Esta sección analiza cómo determinar el tamaño apropiado de la muestra para los estudios que involucran tanto medias como proporciones.

A. Tamaño de la muestra para estimar $\mu_1 - \mu_2$

El tamaño apropiado para la muestra se halla mediante la fórmula (9.14):

Tamaño de la muestra para la diferencia entre medias poblacionales

$$n = \frac{Z^2(\sigma_1^2 + \sigma_2^2)}{(\text{error})^2} \qquad [9.14]$$

La Comisión de Planeación Económica de Texas (*Texas Economic Planning Commission*) pidió a un economista de la Universidad de Texas, en Arlington, que desarrollara un intervalo de confianza del 99% para la diferencia entre la duración promedio del servicio prestado por empleados públicos y el de los trabajadores del sector privado. La comisión desea un ancho de intervalo de tres años. Las muestras piloto produjeron varianzas de 15 y 21, respectivamente. ¿Qué tan grandes deberían tomarse las muestras de cada población?

Debido a que el intervalo es de 3 años, el error es de la mitad de dicha cifra, o 1.5 años. Entonces:

$$n = \frac{(2.58)^2(15 + 21)}{(1.5)^2}$$
$$= 106.5, \text{ o } 107 \text{ mujeres}$$

Por tanto, 107 empleados deberían seleccionarse del sector público y 107 deberían seleccionarse del sector privado para hacer la comparación.

B. Tamaño de la muestra para estimar $\pi_1 - \pi_2$

El tamaño correcto de la muestra para estimar la diferencia entre dos proporciones poblacionales se halla mediante la fórmula (9.15):

Tamaño de la muestra para la diferencia entre proporciones poblacionales

$$n = \frac{Z^2[\pi_1(1 - \pi_1) + \pi_2(1 - \pi_2)]}{(\text{error})^2} \qquad [9.15]$$

Wally Simpleton, el candidato líder en la carrera por la gobernación, desea desarrollar un intervalo de confianza con un ancho de 3 puntos porcentuales y un nivel de confianza del 99% para hallar la diferencia entre la proporción de hombres y de mujeres que están a favor de su candidatura. ¿Qué tan grandes deberían ser las muestras? Una muestra piloto para hombres y mujeres reveló que p_h ¿ 0.40 y $p_m = 0.30$. Si el ancho de intervalo es 0.03, el error es 0.03/2 = 0.015.

$$n = \frac{(2.58)^2[(0.4)(0.6) + (0.3)(0.7)]}{(0.015)^2}$$
$$= 13{,}312 \text{ hombres y } 13{,}312 \text{ mujeres}$$

9.6 Pruebas de hipótesis para dos medias con muestras independientes

Las pruebas de hipótesis para diferencias entre las medias siguen un procedimiento parecido al de los intervalos en que las muestras son o bien 1) independientes o 2) pareadas. Sin embargo, en este caso, a diferencia del caso de la estimación por intervalo, no se está interesado en el tamaño de la diferencia en las medias, sino sólo en si existe o no una diferencia. Por ejemplo, una prueba de dos colas se expresaría como:

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

o el equivalente

$$H_o\text{: } \mu_1 - \mu_2 = 0$$
$$H_A\text{: } \mu_1 - \mu_2 \neq 0$$

Esta sección analiza las pruebas con muestras independientes tanto para muestras grandes como para muestras pequeñas.

A. Pruebas con muestras grandes

La prueba de hipótesis para la diferencia entre medias de dos poblaciones es muy parecida a la de una sola población. Se incluyen los mismos cuatro pasos:

1. Se plantean las hipótesis
2. Se calcula un estadístico Z o t.
3. Se forma la regla de decisión
4. Interpretación y conclusiones.

El valor de Z es

Estadístico de prueba Z para muestras grandes	$$Z = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{s_{\bar{x}_1 - \bar{x}_2}}$$	[9.16]

en donde $s_{\bar{x}_1 - \bar{x}_2}$ es el estimado para el error estándar de las diferencias en las medias muestrales, tal y como se hizo para la estimación de intervalo. Claro que si las varianzas poblacionales son conocidas, debería utilizarse $\sigma_{\bar{x}_1 - \bar{x}_2}$.

Weaver Ridge Golf Course desea ver si el tiempo promedio que requieren los hombres para jugar los 18 hoyos es diferente al de las mujeres. Se mide el tiempo de cincuenta partidos dobles de hombres y 45 de mujeres obteniendo

Hombres	Mujeres
$\bar{X}$ = 3.5 horas	$\bar{X}$ = 4.9 horas
s = 0.9 horas	s = 1.5 horas

Vale la pena recordar de la discusión sobre intervalos de confianza que

$$s_{\bar{x}_1-\bar{x}_2} = \sqrt{\frac{(0.9)^2}{50} + \frac{(1.5)^2}{45}} = 0.257$$

Entonces,

$$Z = \frac{(3.5 - 4.9) - 0}{0.257} = -5.45$$

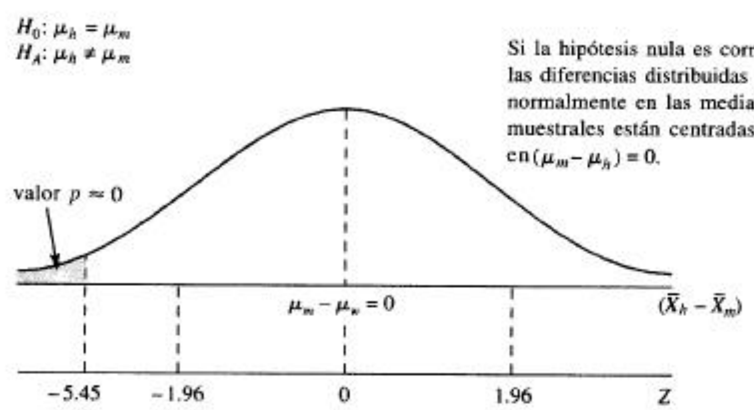
Si $\alpha = 0.05$, el valor crítico de Z es ± 1.96 y la regla de decisión es:

> **Regla de decisión:** "No rechazar si Z está entre ± 1.96. Rechazar si Z es menor que – 1.96 o más que + 1.96."

Como lo muestra la figura 9.3, la hipótesis nula se rechaza. Debido a que la hipótesis nula de igualdad es rechazada y $\bar{X}_m > \bar{X}_h$, la evidencia sugiere que las mujeres toman más tiempo en promedio. Vale la pena notar también que el valor p relacionado con la prueba es virtualmente cero.

Figura 9.3

Prueba de hipótesis para la igualdad de los tiempos del juego de golf



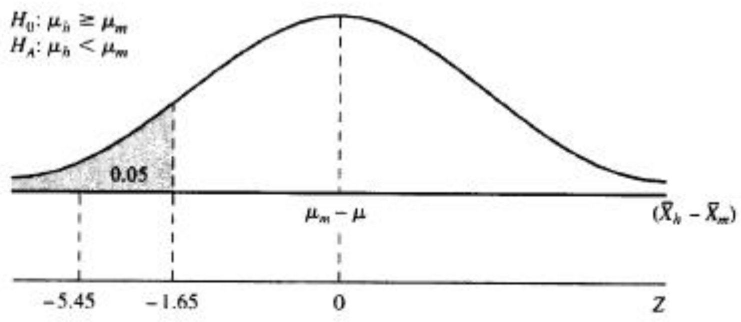
Si con el recorrido del campo de golf se hubiera planteado como hipótesis que los hombres toman menos tiempo ($\mu_h < \mu_m$), las hipótesis se plantearían como

$$H_0\text{: } \mu_h \geq \mu_m$$
$$H_A\text{: } \mu_h < \mu_m$$

y se realizaría una prueba de cola a la izquierda como se observa en la figura 9.4. Si $\alpha = 0.05$ se mantiene, la prueba de cola a la izquierda requiere un valor crítico de Z de –1.65. El valor Z de –5.45 no cambia y de nuevo cae en la zona de rechazo.

Figura 9.4

Prueba de una sola cola para los tiempos promedios de golf



Ejemplo 9.6

En el ejemplo 9.1 anterior, los gerentes del Charles Schwab construyeron un estimado de intervalo del 99% para la diferencia entre los niveles de competencia promedio de los dos grupos de empleados. El resultado fue $-8.34 \leq (\mu_1 - \mu_2) \leq 4.40$. Se supone que querían probar la hipótesis de que las competencias promedio eran iguales, entonces

Solución

Las hipótesis son:

$$H_o: \mu_1 = \mu_2$$
$$H_A: \mu_1 \neq \mu_2$$

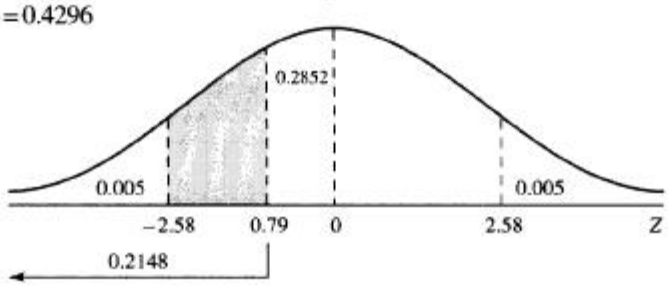
Dados los datos del ejemplo 9.1,

$$Z = \frac{(76 - 77.97) - 0}{2.47} = -0.79$$

Si $\alpha = 0.01$, la regla de decisión es

Regla de decisión: "No rechazar si Z está entre ± 2.58. De otro modo rechazar".

Como lo muestra la figura, Z = –0.79, la hipótesis nula no es rechazada. Además, el valor p es $2(0.5 - 0.2852) = 0.4296$



Interpretación

La evidencia sugiere que no hay diferencia en la competencia promedio de los grupos de empleados entrenados con los dos programas. Esta posición se confirma por el hecho de que el intervalo calculado en el ejemplo 9.1 contenía cero, indicando que no hay diferencia.

La impresión en Minitab para las pruebas de Schwab se observan en la pantalla 9.1. Haga clic en **Stat > Basic Statistics > 2-sample t**. Esto dará los resultados tanto del estimado de intervalo como de la prueba de hipótesis. Vale la pena notar el intervalo de confianza del 99% así como el valor t y el valor p.

Pantalla 9.1

Two sample T-test and confidence interval (Prueba T con dos muestras e intervalo de confianza)

```
Twosample T for Prog 1 vs Prog 2 (Prueba T con dos muestras para Prog 1 contra Prog 2)

          N       Mean                StdDev       SE Mean
                 (Media)   (Desviación estándar)  (Media SE)
Prog 1   45       76.0                  13.5          2.0
Prog 2   40      77.97                  9.05          1.4

99% C.I. for mu Prog 1 - mu Prog 2: (-8.4,  4.6)                   DF = Grados
[99% I.C. para mu Prog 1 - mu Prog 2: (-8.4,  4.6)]                     de libertad
T-test mu Prog 1 = mu Prog 2 (vs not =): T= -0.78  P=0.44  DF=77
[Prueba T para mu Prog 1 = mu Prog 2 (contra no =): T= -0.78  P=0.44  DF=77]
```

B. Pruebas con muestras pequeñas: La distribución t

De la misma manera que en el caso de los intervalos de confianza, las pruebas que involucran muestras pequeñas dependen de si las varianzas poblacionales pueden asumirse como iguales, o si permiten datos mancomunados. Si se permite la suposición de varianzas iguales, la prueba t se vuelve

Prueba de hipótesis con muestras pequeñas cuando $\sigma_1^2 = \sigma_2^2$ (desconocidas)	$t = \dfrac{(\overline{X}_1 - \overline{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\dfrac{s_p^2}{n_1} + \dfrac{s_p^2}{n_2}}}$	[9.17]

Regresando al ejemplo 9.2, un estimado de intervalo del 98% de la diferencia en los salarios promedio de los trabajadores en Atlanta y Newport News se calculó con base en

Atlanta	Newport New
$n = 23$	$n = 19$
$\overline{X}$ = US\$17.53	$\overline{X}$ = US\$15.50
$s^2 = 92.10$	$s^2 = 87.10$

El intervalo era de $-5.09 \le (\mu_1 - \mu_2) \le 9.15$. Si en lugar de un estimado por intervalo se hubiera querido realizar una prueba de hipótesis para medias iguales, entonces se tendría

$$H_o\text{: } \mu_A = \mu_N$$
$$H_A\text{: } \mu_A \neq \mu_N$$

Aplicando estos datos a la fórmula (9.17), se tiene que

$$t = \frac{(17.53 - 15.5) - 0}{\sqrt{\dfrac{89.85}{23} + \dfrac{89.85}{19}}} = 0.69$$

Dado que α es 2% y que hay $23 + 19 - 2 = 40$ grados de libertad, el valor crítico de t es ± 2.423. La hipótesis nula por tanto no se rechaza. Parece que no hay diferencia en el salario promedio. Esta conclusión se confirma por el hecho de que el intervalo contenía cero.

Si no se hace la suposición de varianzas iguales, se tiene que

Prueba con muestras pequeñas con varianzas desiguales	$t = \dfrac{(\overline{X}_1 - \overline{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}}}$	[9.18]

t se compara con el valor crítico de t con base en los grados de libertad determinados por la fórmula (9.7). En el ejemplo 9.3, un intervalo del 98% para la diferencia en la durabilidad promedio de los dos tipos de amortiguadores de caucho para coches de bebé, se estimó que era

$$0.5 \text{ semanas } \le (\mu_1 - \mu_2) \le 7.1 \text{ semanas}$$

Los resultados se basaron en los datos

$$\overline{X}_1 = 11.3 \text{ semanas} \quad \overline{X} = 7.5 \text{ semanas}$$
$$n_1 = 13 \quad n_2 = 10$$
$$s_1 = 3.5 \text{ semanas} \quad s_2 = 2.7 \text{ semanas}$$

Una prueba de hipótesis para igualdad de medias sería

$$H_o\colon \mu_1 = \mu_2$$
$$H_A\colon \mu_1 \neq \mu_2$$

t es

$$t = \frac{(11.3 - 7.5) - 0}{\sqrt{\frac{(3.5)^2}{13} + \frac{(2.7)^2}{10}}} = 2.94$$

Si $\alpha = 0.02$ y los grados de libertad son 20 con base en la fórmula (9.7), la regla de decisión es

Regla de decisión: "No rechazar si t está entre ±2.528. De otro modo rechazar".

Debido a que $2.94 > 2.528$, la hipótesis nula de igualdad es rechazada. Dado $\overline{X}_1 > \overline{X}_2$, la evidencia sugiere que el tipo 1 de amortiguador de caucho para coche de bebé presenta mayor durabilidad. De nuevo esta afirmación se confirma con el hecho de que el intervalo del ejemplo 9.3 no contenía cero.

9.7 Pruebas de hipótesis con datos por pareados

Las muestras por pares para realizar una prueba de hipótesis ofrecen las mismas ventajas que en la construcción de intervalos de confianza. Se experimenta menos error de muestreo debido a la reducción en la variación entre observaciones porque se hace la correspondencia tan cerca como sea posible.

t es

Prueba de hipótesis para muestras pareadas	$t = \frac{\bar{d} - (\mu_1 - \mu_2)}{\frac{s_d}{\sqrt{n}}}$	[9.19]

en donde $\overline{d}$ es la media de las diferencias en las observaciones pareadas y s_d es el error estándar de dichas diferencias y se halla utilizando la fórmula (9.10).

En el ejemplo 9.4, Vicki Peplow preparó un estimado de intervalo del 95% para la diferencia en costos para procedimientos idénticos en los dos hospitales. El resultado fue

$$\text{US\$}{-146.33} \leq (\mu_1 - \mu_2) \leq \text{US\$}28.47$$

con base en $n = 15$, $\Sigma d_i = -884$, y $\Sigma d_i 2 = 400{,}716$. Si la señora Peplow fuera a probar una hipótesis de igualdad, hallaría

$$H_o\colon \mu_1 = \mu_2$$
$$H_A\colon \mu_1 \neq \mu_2$$

y

$$t = \frac{-58.93 - 0}{\frac{157.8}{\sqrt{15}}} = -1.44$$

Dado $t_{.05,14} = \pm\, 2{,}145$, la regla de decisión es:

Regla de decisión: "No rechazar si t está entre $\pm$ 2.145, de lo contrario rechazar".

Esto termina en un no rechazo de la hipótesis nula.

Ejercicios de la sección

22. Muestras de tamaños 50 y 60 revelan medias de 512 y 587, y desviaciones estándar de 125 y 145 respectivamente. A un nivel del 2%, pruebe la hipótesis de que $m_1 = m_2$.

23. A un nivel del 1%, pruebe la igualdad de las medias si muestras de tamaño 10 y 8 dan medias de 36 y 49 y desviaciones estándar de 12 y 18, respectivamente. Se asume que las varianzas no son iguales.

24. Responda el problema anterior asumiendo que las varianzas son iguales.

25. Muestras pareadas de tamaño 81 dan una media de las diferencias de 36.5 y una desviación estándar de las diferencias de 29.1. Pruebe la igualdad de las medias. Sea $a = 0.01$.

26. Pruebe H_0: m_1 £ m_2 si los tamaños de las muestras de 64 y 81 producen medias de 65.2 y 58.6 y desviaciones estándar de 21.2 y 25.3. Sea $a = 5\%$.

27. Pruebe H_0: m_1 ³ m_2 si dos muestras de tamaño 100 producen medias de 2.3 y 3.1 con desviaciones estándar de 0.26 y 0.31. Sea $a = 1\%$.

28. Las muestras pareadas de tamaño 25 reportaron una diferencia media de 45.2 y una desviación estándar de las diferencias de 21.6. Pruebe la igualdad de las medias a un nivel del 5%.

9.8 Una prueba para la diferencia entre dos proporciones

Con frecuencia, en el mundo de los negocios surgen problemas que requieren de la comparación entre dos proporciones de poblaciones diferentes: la proporción de defectos producidos por un método, comparado con los producidos por un segundo método; o la proporción de deudas vencidas en una empresa comparada con las de otra empresa.

La prueba para medir estas proporciones es la siguiente:

Prueba para la diferencia de dos proporciones

$$Z = \frac{(p_1 - p_2) - (\pi_1 - \pi_2)}{s_{p_1 - p_2}} \qquad [9.20]$$

en donde p_1 y p_2 son proporciones de éxitos en las muestras y s_{p1-p2} es la desviación estándar de tales diferencias. Se calcula con la fórmula (9.12) que se utilizó en la construcción de los intervalos de confianza.

Para ilustrar, un minorista desea probar la hipótesis de que la proporción de sus clientes masculinos, quienes compran a crédito, es igual a la proporción de mujeres que utilizan el crédito. Él selecciona 100 clientes hombres y encuentra que 57 compraron a crédito mientras que 52 de las 110 mujeres lo hicieron.

Las hipótesis son

$$H_0: \pi_h = \pi_m$$
$$H_A: \pi_h \neq \pi_m$$

A un nivel del 1% la regla de decisión es:

Regla de decisión: "No rechazar si $-2.58 \leq Z \leq 2.58$. De lo contrario rechazar".

$$s_{p_1-p_2} = \sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}$$

$$= \sqrt{\frac{(0.57)(0.43)}{100} + \frac{(0.473)(0.527)}{110}}$$

$$= 0.069$$

Entonces,

$$Z = \frac{(0.57 - 0.473) - 0}{0.069}$$

$$= 1.41$$

Debido a que t está entre $\pm$ 2.58, la hipótesis nula no se rechaza. El minorista no puede concluir a un nivel del 1% que las proporciones de hombres y mujeres que compran a crédito difieren.

Ejemplo 9.7

Recientemente Johnson Manufacturing ha experimentado un incremento en el número de unidades defectuosas. El supervisor de producción considera que el turno de la noche produce una proporción más elevada de defectos que los del turno de día: $\pi_N > \pi_D$. Para comparar la proporción de defectos, se toma una muestra de 500 unidades de la producción del turno de día y revela 14 defectos. Una muestra de 700 unidades del turno de la noche muestra 22 defectos. Si una proporción más grande de defectos se origina en la producción nocturna, el supervisor pretende instituir un programa de capacitación para que los trabajadores mejoren sus destrezas laborales. ¿A un nivel del 5%, debería implementarse el programa?

Solución

Debido a que se pensaba que los trabajadores nocturnos producían una proporción de defectos más alta, las hipótesis son

$$H_o: \pi_N \leq \pi_D$$
$$H_A: \pi_N > \pi_D$$

y se requiere una prueba de cola a la derecha. La regla de decisión es:

Regla de decisión: "No rechazar si $Z \leq 1.65$. Rechazar si $Z > 1.65$".

La proporción de defectos del turno de la noche es $p_N = 22/700 = 0.031$ y la del turno diurno es $p_D = 14/500 = 0.028$; entonces

$$s_{p_1-p_2} = \sqrt{\frac{(0.031)(0.969)}{700} + \frac{(0.028)(0.972)}{500}}$$

$$= 0.0099$$

y

$$Z = \frac{(0.031 - 0.028) - 0}{0.0099}$$
$$= 0.303$$

Interpretación

Debido a que Z es menor que 1.65, la hipótesis nula $\pi_N \leq \pi_D$ no se rechaza. La evidencia no es lo suficientemente fuerte como para concluir que la tasa de defectos de los trabajadores del turno de la noche sea mayor que la del turno de día. No debe instituirse el programa de capacitación.

Ejercicios de la sección

29. Muestras de tamaños 120 y 150 produjeron proporciones de 0.69 y 0.73. Pruebe la igualdad de las proporciones de la población al nivel del 5%.

30. Dos muestras de tamaño 500 cada una se utilizaron para probar la hipótesis de que H_0: $\mu_1 \leq \mu_2$. Las proporciones muestrales son 14% y 11%. A un nivel del 10%, ¿cuál es su conclusión?

31. Muestras de tamaños 200 y 250 revelan proporciones muestrales de 21% y 26%. Pruebe la hipótesis de que H_0: $\pi_1 \geq \pi_2$. Fije α al 1%.

9.9 Comparación de la varianza de dos poblaciones normales

Varias de las pruebas estadísticas discutidas anteriormente partían de la suposición de varianzas poblacionales iguales. En su momento se pidió aceptar a ciegas esta igualdad, con la promesa de que posteriormente se demostraría cómo probarlo. Pues el momento ha llegado: esta sección demuestra cómo determinar si la suposición de varianzas iguales es razonable. Esta prueba se basa en la distribución F, la cual fue denominada así en 1924 en honor a Sir Ronald A. Fisher (1890-1962).

Cuando se comparan las varianzas de dos poblaciones, se toma una muestra de cada población. Las varianzas de la muestra sirven como estimados de sus varianzas poblacionales respectivas. Una distribución F se forma por la razón de estas dos varianzas muestrales. La razón F es

Razón F utilizada para comparar dos varianzas poblacionales

$$F = \frac{s_L^2}{s_s^2} \qquad [9.21]$$

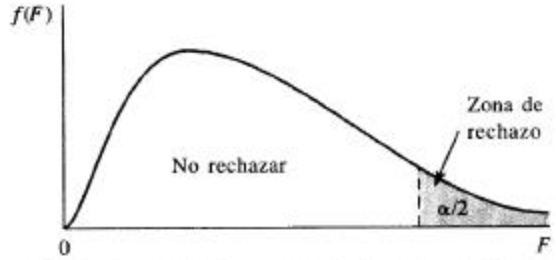
en donde s_L^2 es la más grande de las dos varianzas muestrales y s_s^2 es la más pequeña de las dos varianzas muestrales.

Se asume que la varianza de la segunda muestra excede la de la primera. En ese caso, la razón F es $F = s_2^2/s_1^2$. Entre más exceda s_2^2 a s_1^2, menos probable es que $\sigma_1^2 = \sigma_2^2$ y más grande será la razón F. Por tanto, un valor F grande conllevará a un rechazo de la hipótesis nula H_0: $\sigma_1^2 = \sigma_2^2$.

Manipular la razón F para garantizar que la varianza de la muestra más alta se coloque en el numerador asegurará que el valor F siempre sea mayor que 1. Vale la pena destacar en la figura 9.5 que la distribución F no es simétrica y está limitada por cero en el extremo inferior.

Debido a que la razón F está restringida para exceder siempre de 1, resulta sólo una zona de rechazo en la cola derecha. Está prohibida toda zona de rechazo que hubiera aparecido de otro modo en la cola izquierda. Por tanto, sólo la mitad del área bajo la distribución es accesible como zona de rechazo y, es necesario dividir por 2 el valor α seleccionado e identificar una zona de rechazo única en la cola derecha, con un área igual a $\alpha/2$ tal como aparece en la figura 9.5.

Figura 9.5
Distribución F



El valor α Cuando se controla la razón F para asegurar que $F > 1$, se realiza la prueba de dos colas de la hipótesis H_0: $\sigma_1^2 = \sigma_2^2$ como si fuera una prueba de una cola. Por tanto, es necesario dividir por 2 el valor α.

Se supone que se escogió originalmente un valor α del 10%. Dividir por 2, $\alpha/2 = 0.05$, requiere que se consulte la sección de la tabla G (apéndice III) que pertenece a un α del 5%. Además, la distribución F tiene dos grados de libertad: uno para el numerador, el cual es igual a $n_1 - 1$ y otro grado para el denominador, el cual es igual a $n_2 - 1$, en donde n_1 y n_2 son los tamaños de las muestras en el numerador y en el denominador, respectivamente. La notación convencional para el valor crítico de F es $F_{\alpha/2, n_1-1, n_2-1}$.

Un consultor gerencial desea probar una hipótesis respecto a dos medias poblacionales. Sin embargo, antes de hacerlo debe decidir si hay alguna evidencia que sugiera que las varianzas poblacionales son iguales. Al recolectar sus datos, el consultor encuentra que:

	Muestra 1	Muestra 2
Tamaño muestral	10	10
Desviación estándar	12.2	15.4
Varianza	148.84	237.16

El desea probar

$$H_o: \sigma_1^2 = \sigma_2^2$$
$$H_A: \sigma_1^2 \neq \sigma_2^2$$

Utilizando la fórmula (9.21), y asegurándose de que la varianza más alta esté en el numerador, se encuentra que

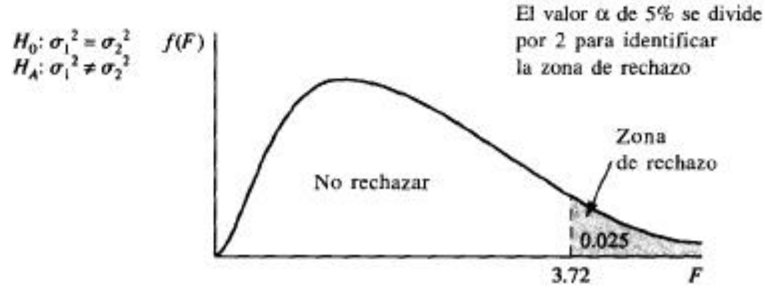
$$\begin{aligned} F &= \frac{s_1^2}{s_2^2} \\ &= \frac{(15.4)^2}{(12.2)^2} \\ &= 1.59 \end{aligned}$$

Si α se fija al 5%, $F_{.05/2,9,9} = 4.03$. Esto se encuentra en la tabla G (apéndice III) localizando dicha sección de la tabla que tiene que ver con un α de $0.05/2 = 0.025$. Se mueve hacia la fila superior para hallar 9 grados de libertad para el numerador y luego se baja en la columna 9 grados de libertad para hallar el denominador. Allí se encontró el valor 4.03. Como se muestra en la figura 9.6.

La **regla de decisión** es: "No rechazar si $F \leq 4.03$. Rechazar si $F > 4.03$".

Debido a que $F = 1.59 < 4.03$, la hipótesis nula H_0: $\sigma_1^2 = \sigma_2^2$ no se rechaza. El consultor puede proceder con la prueba de hipótesis respecto a las medias poblacionales bajo la suposición de que las varianzas son iguales.

Figura 9.6
Distribución F para la igualdad de varianzas



Problemas resueltos

1. **Ética del Trabajo de los Yuppies** *(The Yuppies' Work Ethic)* Una emisión del mes de abril de 1991 de la revista *Fortune* contaba una historia sobre los adictos al trabajo nacidos después de la Segunda Guerra Mundial, entre los 25 y 43 años de edad, que tenían trabajos en el área administrativa. El artículo comparaba la vida laboral de estos ejecutivos jóvenes que se habían ubicado en la vía rápida corporativa, con las de los trabajadores quienes dedicaban menos tiempo a su trabajo. A pesar de que quienes estaban orientados hacia el éxito reportaron con mayor frecuencia 70, 80 o incluso 90 horas semanales en el trabajo, aproximadamente 60 horas era lo típico. Los datos se recolectaron de las entrevistas hechas a empleados corporativos. Sea el grupo 1 los que se ubican en la vía rápida y el grupo 2 los que pasan menos tiempo en el trabajo. Se asume que las entrevistas revelaron las siguientes estadísticas respecto a los horarios de trabajo semanales.

Grupo 1	Grupo 2
$\bar{X}_1 = 62.5$ horas	$\bar{X}_2 = 39.7$ horas
$s_1 = 23.7$ horas	$s_2 = 8.9$ horas
$n_1 = 175$	$n_2 = 168$

Haga e interprete un intervalo del 90% para la diferencia en las horas promedio de trabajo y pruebe la hipótesis de medias iguales a un nivel del 10%.

Solución

Dado

$$s_{\bar{x}_1-\bar{x}_2} = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} = \sqrt{\frac{(23.7)^2}{175} + \frac{(8.9)^2}{168}}$$
$$= 1.92$$

Entonces,

$$\begin{aligned}\text{I.C. para } (\mu_1 - \mu_2) &= (\bar{X}_1 - \bar{X}_2) \pm Z s_{\bar{x}_1-\bar{x}_2} \\ &= (62.5 - 39.7) \pm (1.65)(1.92) \\ 19.63 \leq (\mu_1 - \mu_2) &\leq 25.97\end{aligned}$$

Usted puede estar 90% seguro de que quienes van en la vía rápida trabajan un promedio de 19.63 horas a 25.97 horas más.

La prueba de hipótesis es

$$H_o\text{: } \mu_1 = \mu_2$$
$$H_A\text{: } \mu_1 \neq \mu_2$$
$$Z = \frac{(62.5 - 39.7) - 0}{1.92} = 11.88$$

La **regla de decisión** es: "No rechazar si Z está entre ± 1.65. De lo contrario rechazar".

Debido a que Z = 11.88 > 1.65, la hipótesis nula se rechaza.

2. **Inflación y poder del mercado** Muchos estudios económicos se concentran en las industrias en las cuales una buena parte del poder de mercado se concentra en manos de sólo unas pocas empresas. Se teme que las empresas poderosas de las industrias de tan alta concentración dominen el mercado para su propio beneficio. Fueron pareadas empresas en nueve industrias con empresas, en un número igual de industrias, en las cuales el poder económico estaba más disperso. Se equipararon las industrias de cada grupo con respecto a la competencia extranjera, estructuras de costos y otros factores que pueden afectar los precios de la industria. Los incrementos promedio en precios en los porcentajes de cada industria aparecen en la siguiente tabla. A un nivel del 10% ¿parece que las industrias concentradas presentan presiones inflacionarias más pronunciadas que las industrias menos concentradas? Haga el intervalo apropiado y pruebe la hipótesis apropiada.

Pares de industrias	Industrias concentradas (%)	Industrias menos concentradas (%)	d_i	d_i^2
1	3.7	3.2	0.5	0.25
2	4.1	3.7	0.4	0.16
3	2.1	2.6	-0.5	0.25
4	-0.9	0.1	-1.0	1.00
5	4.6	4.1	0.5	0.25
6	5.2	4.8	0.4	0.16
7	6.7	5.2	1.5	2.25
8	3.8	3.9	-0.1	0.01
9	4.9	4.6	0.3	0.09
			2.0	4.42

Solución

$$\bar{d} = \frac{\Sigma d_i}{n} = \frac{2}{9} = 0.22$$

$$\begin{aligned}s_d &= \sqrt{\frac{\Sigma d_i^2 - n\bar{d}^2}{n-1}} = \sqrt{\frac{4.42 - 9(0.22)^2}{8}} \\ &= 0.706\end{aligned}$$

$$\text{I.C. para } (\mu_1 - \mu_2) = \bar{d} \pm t\frac{s_d}{\sqrt{n}}$$
$$= 0.22 \pm (1.860)\frac{0.706}{\sqrt{9}}$$
$$-0.218 \leq \mu_d \leq 0.658$$

Se está 90% seguro de que el incremento promedio en precios de las industrias menos concentradas es de 0.658% más bajo a 0.218% más alto.

La prueba de hipótesis es:

$$H_o\colon \mu_1 = \mu_2$$
$$H_A\colon \mu_1 \neq \mu_2$$
$$t = \frac{\bar{d} - (\mu_1 - \mu_2)}{\frac{s_d}{\sqrt{n}}} = \frac{0.22 - 0}{\frac{0.706}{\sqrt{9}}}$$
$$= 0.935$$

La **regla de decisión** es: "No rechazar si t está entre $\pm$ 1.860. De lo contrario rechazar".

Debido a que $0.935 < 1.860$, no se rechaza la hipótesis nula.

3. Pruebas t con muestras pequeñas Una compañía de perforación prueba dos brocas de barrena calando a un máximo de 112 pies y registrando el número de horas que tomó el procedimiento. La primera broca se utilizó en 12 casos, resultando un tiempo promedio de $\bar{X}_1 = 27.3$ horas y $s_1 = 8.7$ horas. Se excavaron 10 pozos con una segunda broca, produciendo un $\bar{X}_2 = 31.7$ horas y $s_2 = 8.3$ horas.

¿Parece que una broca es más efectiva que la otra? Sea $\alpha = 0.10$. Ninguna evidencia sugiere que las varianzas son iguales. Responda construyendo el intervalo apropiado.

Solución

$$\text{g.l.} = \frac{[s_1^2/n_1 + s_2^2/n_2]^2}{\frac{(s_1^2/n_1)^2}{(n_1 - 1)} + \frac{(s_2^2/n_2)^2}{(n_2 - 1)}}$$
$$= \frac{[(8.7)^2/12 + (8.3)^2/10]^2}{\frac{[(8.7)^2/12]^2}{11} + \frac{[(8.3)^2/10]^2}{9}} = \frac{174.15}{8.89}$$
$$= 19.59$$
$$\approx 19$$

Con g.l. = 19, y $\alpha = 0.10$, $t' = 1.729$.

$$\text{I.C. para } (\mu_1 - \mu_2) = (\bar{X}_1 - \bar{X}_2) \pm t'\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$
$$= (27.3 - 31.7) \pm (1.729)\sqrt{\frac{(8.7)^2}{12} + \frac{(8.3)^2}{10}}$$
$$-10.68 \leq (\mu_1 - \mu_2) \leq 1.88$$

Se puede estar 90% seguro de que la perforación con la broca 1 se toma entre 1.88 horas más y 10.68 horas menos.

La prueba de hipótesis es:

$$H_o: \mu_1 = \mu_2$$
$$H_A: \mu_1 \neq \mu_2$$
$$t = \frac{(27.3 - 31.7) - 0}{\sqrt{\frac{(8.7)^2}{12} + \frac{(8.3)^2}{10}}}$$
$$= -1.211$$

La **regla de decisión** es: "No rechazar si t está entre ±1.729. De lo contrario rechazar".

Debido a que $t = -1.211 > -1.729$, no se rechaza la hipótesis nula.

Todos los pozos fueron excavados con el mismo equipo de perforación y en el mismo suelo. Si por este motivo o por cualquier otro que pueda mencionarse, la compañía de perforación tendría varianzas iguales, ¿cómo se diferenciaría en parte la prueba? Varianzas iguales permiten calcular:

$$s_p^2 = \frac{s_1^2(n_1 - 1) + s_2^2(n_2 - 1)}{n_1 + n_2 - 2}$$
$$= \frac{(8.7)^2(12 - 1) + (8.3)^2(10 - 1)}{12 + 10 - 2}$$
$$= 72.63 \text{ horas al cuadrado}$$

Con grados de libertad iguales a $n_1 + n_2 - 2 = 12 - 10 - 2 = 20$, $t_{0.10,20} = 1.725$. Entonces

$$\text{I.C. para } (\mu_1 - \mu_2) = (\bar{X}_1 - \bar{X}_2) \pm t\sqrt{\frac{s_p^2}{n_1} + \frac{s_p^2}{n_2}}$$
$$= (27.3 - 31.7) \pm (1.725)\sqrt{\frac{72.63}{12} + \frac{72.63}{10}}$$
$$-10.69 \leq (\mu_1 - \mu_2) \leq 1.89$$

4. **The Credit Crunch** Un estudio en *Retail Management* reveló que 131 de las 468 mujeres que efectuaron compras al por menor lo hicieron utilizando una tarjeta de crédito en particular, mientras que 57 de los 237 hombres utilizaron la misma tarjeta. ¿Existe evidencia que sugiera una diferencia en la proporción de mujeres y hombres que utilicen esa tarjeta? Sea $\alpha = 0.05$. Construya el intervalo y pruebe la hipótesis.

Solución

$$s_{p_1-p_2} = \sqrt{\frac{p_1(1 - p_1)}{n_1} + \frac{p_2(1 - p_2)}{n_2}}$$
$$= \sqrt{\frac{(0.28)(0.72)}{468} + \frac{(0.24)(0.76)}{237}}$$
$$= 0.035$$
$$\text{I.C. para } (\pi_m - \pi_h) = (0.28 - 0.24) \pm (1.96)(0.035)$$
$$-0.029 \leq (\pi_m - \pi_h) \leq 0.109$$

Parece que no hay diferencia en la proporción de hombres y de mujeres que utilizan el crédito.

La prueba de hipótesis es:

$$H_o: \pi_m = \pi_h$$
$$H_A: \pi_m \neq \pi_h$$
$$Z = \frac{(0.28 - 0.24) - 0}{0.035} = 1.14$$

La **regla de decisión** es: "No rechazar si Z está entre ± 1.96. De lo contrario rechazar".

Debido a que Z = 1.14, no se rechaza la hipótesis nula.

Lista de fórmulas

[9.1]	$(\bar{X}_1 - \bar{X}_2) \pm Z\sigma_{\bar{x}_1-\bar{x}_2}$	Intervalo de confianza para la diferencia entre dos medias poblacionales – muestras grandes
[9.2]	$\sigma_{\bar{x}_1-\bar{x}_2} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$	Error estándar de las diferencias entre medias muestrales
[9.3]	$s_{\bar{x}_1-\bar{x}_2} = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$	Estimación del error estándar de la diferencia entre medias muestrales
[9.4]	$(\bar{X}_1 - \bar{X}_2) \pm Zs_{\bar{x}_1-\bar{x}_2}$	Intervalo de confianza cuando las varianzas poblacionales son desconocidas
[9.5]	$s_p^2 = \frac{s_1^2(n_1 - 1) + s_2^2(n_2 - 1)}{n_1 + n_2 - 2}$	Estimado de la varianza mancomunada común a ambas poblaciones
[9.6]	$(\bar{X}_1 - \bar{X}_2) \pm t\sqrt{\frac{s_p^2}{n_1} + \frac{s_p^2}{n_2}}$	Intervalo para la diferencia entre las medias cuando $\sigma_1^2 = \sigma_2^2$ (desconocidas)
[9.7]	$\frac{(s_1^2/n_1 + s_2^2/n_2)^2}{(s_1^2/n_1)^2/(n_1 - 1) + (s_2^2/n_2)^2/(n_2 - 1)}$	Grados de libertad cuando las varianzas poblacionales no son iguales
[9.8]	$(\bar{X}_1 - \bar{X}_2) \pm t'\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$	Intervalo para la diferencia entre medias poblacionales cuando $\sigma_1^2 \neq \sigma_2^2$ (desconocidas)
[9.9]	$\bar{d} = \frac{\Sigma d_i}{n}$	Media de las diferencias; observaciones pareadas
[9.10]	$s_d = \sqrt{\frac{\Sigma d_i^2 - n\bar{d}^2}{n - 1}}$	Desviación estándar de las diferencias en el caso de observaciones pareadas
[9.11]	$\bar{d} \pm t\frac{s_d}{\sqrt{n}}$	Intervalo para la diferencia de medias; observaciones pareadas

[9.12] $$\sqrt{\frac{p_1(1-p_1)}{n_1}+\frac{p_2(1-p_2)}{n_2}}$$ Error estándar para la diferencia de proporciones muestrales

[9.13] $$(p_1-p_2)\pm(Z)s_{p_1-p_2}$$ Intervalo para la diferencia de proporciones poblacionales

[9.14] $$\frac{Z^2(\sigma_1^2+\sigma_2^2)}{(\text{error})^2}$$ Tamaño muestral para la diferencia de medias poblacionales

[9.15] $$\frac{Z^2[\pi_1(1-\pi_1)+\pi_2(1-\pi_2)]}{(\text{error})^2}$$ Tamaño muestral para la diferencia de proporciones poblacionales

[9.16] $$Z=\frac{(\bar{X}_1-\bar{X}_2)-(\mu_1-\mu_2)}{s_{\bar{x}_1-\bar{x}_2}}$$ Prueba de hipótesis para la diferencia de medias – muestras grandes

[9.17] $$t=\frac{(\bar{X}_1-\bar{X}_2)-(\mu_1-\mu_2)}{\sqrt{\frac{s_p^2}{n_1}+\frac{s_p^2}{n_2}}}$$ Prueba de hipótesis con muestras pequeñas cuando $\sigma_1^2=\sigma_2^2$ (desconocidas)

[9.18] $$t=\frac{(\bar{X}_1-\bar{X}_2)-(\mu_1-\mu_2)}{\sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}}}$$ Prueba con muestras pequeñas con varianzas poblacionales desiguales

[9.19] $$t=\frac{\bar{d}-(\mu_1-\mu_2)}{\frac{s_d}{\sqrt{n}}}$$ Prueba de hipótesis para observaciones pareadas

[9.20] $$Z=\frac{(p_1-p_2)-(\pi_1-\pi_2)}{s_{p_1-p_2}}$$ Prueba para la diferencia de dos proporciones

[9.21] $$F=\frac{s_L^2}{s_s^2}$$ La razón F utilizada para comparar dos varianzas poblacionales

Ejercicios del capítulo

32. Un contador de una gran corporación en el centro de Estados Unidos debe decidir si seleccionar AT&T o Sprint para manejar el servicio telefónico de llamadas a larga distancia de la empresa. A continuación se reportan los datos recolectados para muchas llamadas utilizando ambos servicios

	AT&T	Sprint
Número de llamadas	145	102
Costo promedio	US$4.07	US$3.89
Desviación estándar	US$0.97	US$0.85

¿Qué revela un intervalo del 95% sobre la diferencia en las medias poblacionales?

33. The Metro Pet Center compara los costos telefónicos utilizando dos programas de facturación diferentes. Con el primer programa, Metro encontró que sobre 100 semanas la facturación promedio fue de US$32.40, con s = US$15.10. Una muestra de 150 semanas utilizando el segundo programa dio una media de US$47.30, con s= US$13.20. Calcule e interprete el intervalo de confianza del 95% para la diferencia entre las medias de los dos programas.

34. En un artículo sobre viajes comerciales, *U.S. News & World Report* afirmó que el costo promedio en una cadena hotelera a nivel nacional era de US$45.12 por noche, y el de una segunda cadena era de US$42.62 por noche. Se asume que estos estadísticos se basan en muestras de 82 y 97 respectivamente, y que las varianzas poblacionales de cada cadena se sabe que son 9.48 y 8.29 dólares al cuadrado, respectivamente. Usted debe determinar cuál cadena de hoteles utilizará su compañía. A un nivel del 1%, ¿qué le indica el intervalo de confianza sobre la decisión? ¿Hace alguna diferencia cuál cadena utiliza?

35. Se asume que los tamaños muestrales en el problema anterior son $n_1 = 182$ y $n_2 = 197$:

a. Sin solucionar el problema, explique lo que le sucedería al intervalo con estas muestras más grandes. ¿Por qué ocurre?

b. Solucione el problema con estas muestras más grandes y observe si usted estaba en lo correcto.

36. Una emisión reciente de *Business Week* discutía los esfuerzos realizados por una importante empresa de autos para determinar si cierto tipo de vehículo soportaba el uso y desgaste del uso diario, más que un segundo tipo de vehículo. Un especialista en finanzas que se acaba de graduar de una universidad local fue contratado para determinar si existe alguna diferencia en la vida promedio. Ella recolecta datos sobre el número promedio de meses en los que un vehículo está en servicio, antes que sea necesaria la primera reparación importante, y descubre la siguiente información: vehículo 1: $\overline{X}_1 = 27.3$ meses, $s_1 = 7.8$ meses, y $n_1 = 82$ vehículos; vehículo 2: $\overline{X}_2 = 33.3$ meses, $s_2 = 10.4$ meses, y $n_2 = 73$ vehículos. A un nivel del 2%, construya el intervalo de confianza e interprete los resultados.

37. El profesor James desea estimar la diferencia en el tiempo promedio que requieren dos organizaciones para decidir sobre las solicitudes de becas presentadas a ellas. Si 14 solicitudes de becas a la Fundación Nacional de Ciencias (*National Science Foundation- NSF*) tomaron en promedio 45.7 semanas con una desviación estándar de 12.6 semanas, y 12 solicitudes a Servicios Humanos y de Salud (*Health and Human Services*) presentaron valores respectivos de 32.9 y 16.8 semanas, calcule e interprete el intervalo de confianza del 90% para la diferencia entre las medias. Si NSF toma más de 5 semanas más que HHS, James planea presentar sus solicitudes al segundo. ¿Debe hacerlo? Se asume que las varianzas son iguales.

38. Para comparar las medias de los ingresos por ventas diarias, un minorista selecciona un tamaño de muestra de 12 semanas en una tienda, con ingresos promedio de US$125.40 y una desviación estándar de US$34.50, y una muestra de 15 semanas de otra tienda, con una media de US$117.20 y una desviación estándar de US$21.50. Calcule e interprete el intervalo de confianza del 99% para la diferencia entre las medias poblacionales. Sea $\sigma_1^2 = \sigma_2^2$.

39. A un equipo administrativo se pidió solucionar 10 problemas diferentes de control de calidad que se encuentran comúnmente en su trabajo. A un segundo equipo se le pidió solucionar los mismos problemas. Los tiempos de solución en minutos que necesitó cada equipo aparecen a continuación. Calcule e interprete el intervalo del 90% de confianza para la diferencia entre los tiempos promedio de la población que requieren los dos equipos. ¿Qué puede concluir sobre la capacidad de solución de problema relativo de los dos equipos?

Problema	Equipo 1	Equipo 2
1	12	25
2	15	26
3	14	21
4	21	23
5	19	31
6	12	19
7	25	35
8	18	28
9	17	27
10	20	26

40. *The Wall Street Journal* reportó que Ford Motor Company se interesó en los salarios promedio de sus ejecutivos que trabajaban en el exterior en comparación a los que trabajaban en el estado, en el cual quedaba la sede principal. El salario promedio de 87 ejecutivos en el exterior era de US$78,010, con una desviación estándar de US$15,700. El mismo número de ejecutivos que trabajan en el servicio nacional reveló una media y una desviación estándar de US$69,410 y US$10,012 respectivamente. Desarrolle e interprete el intervalo del 97% para las diferencias promedio.

41. Sammy Shopper desea comparar los costos de 10 servicios diferentes ofrecidos por los nuevos concesionarios de autos en esta zona. Aquí se presentan los datos. Calcule e interprete el intervalo de confianza del 90% para la diferencia entre las medias poblacionales. Si la calidad del servicio es la misma, ¿cuál comercializador debería utilizar Sammy? ¿o hay alguna diferencia?

Servicio	Consumidor 1	Consumidor 2
1	US$54	US$36
2	56	35
3	59	34
4	65	39
5	62	37
6	43	32
7	38	31
8	48	30
9	46	29
10	59	45

42. *The Wall Street Journal* reportó que se había aumentado la preocupación respecto al ambiente en el cual el ganado vacuno se mantenía antes del sacrificio. Supuestamente, los ambientes libres de tensiones mejoraban el engorde y la calidad de la carne. Un criador en California del norte incluso anuncia que "lleva a su ganado a una vista espectacular de la costa" antes de prepararlo para llevarlo al puesto de carnes en una tienda local. Se asume que 50 reses de las criadas en este ambiente vacacional ganan en promedio 112 libras con $s = 32.3$ libras durante un período dado. Durante el mismo período, 50 vacas con vista al matadero ganan 105.7 libras en promedio con $s = 28.7$ libras. Calcule e interprete el intervalo del 90%.

43. Hace varias navidades una parte de los renos de Santa Claus se sindicalizaron. Desde entonces Santa Claus se ha preguntado si había diferencia en la productividad promedio de los renos sindicalizados y los no sindicalizados. Una muestra de 150 renos sindicalizados reportaron una producción promedio de 27.3 juguetes por semana por reno, con una desviación estándar de 8.7 juguetes. Una muestra de 132 renos no sindicalizados reveló una media de 29.7 juguetes por semana por reno, con $s = 10.7$. ¿Qué dice un intervalo del 90% a Santa Claus sobre la diferencia en la producción promedio?

44. Veintiséis fondos mutuos, cada uno con US$5,000 invertidos en ellos se seleccionan para hacer una comparación. De los 26 fondos, 12 están orientados hacia los ingresos a corto plazo y produjeron un rendimiento promedio de US$1,098,60 (incluyendo el incremento de capital), con una desviación estándar de US$43.20. Los fondos restantes son de inversión a largo plazo y orientados al desarrollo y generaron una utilidad promedio de US$987.60 (incluyendo los rendimientos de capital), con una desviación estándar de US$53.40.

a. Calcule e interprete el intervalo de confianza del 80% para la diferencia entre los rendimientos promedio poblacionales. No existe motivo para creer que las varianzas son iguales.

b. ¿Qué tamaño muestral es necesario para estar 95% seguro de que el error no excede de US$10.00?

45. The Baldwin Piano Company ha argumentado durante mucho tiempo que su método para enseñar a la gente a tocar el piano es superior al de la competencia. Para estimar cualquier diferencia en el tiempo que se toma aprender las técnicas básicas, usted, como nuevo analista de Baldwin, selecciona 100 estudiantes que utilizaron su método y encuentra que el tiempo promedio fue de 149 horas con $s = 37.7$ horas.

Las estadísticas correspondientes a 130 estudiantes que utilizaron el método de la competencia fue de 186 y 42.2.

a. A un nivel de confianza del 99%, ¿qué puede concluir sobre la diferencia en los tiempos promedio de aprendizaje? ¿Parece que su compañía ofrece un método mejor?

b. Qué tamaño de muestra se necesita para estar 99% seguro de que el error no excede de 5 horas?

46. Mientras que laboraba como interno practicante para una importante empresa aseguradora, un miembro administrativo de la universidad local realizó un estudio para medir la expectativa de vida promedio de los alcohólicos en oposición a quienes no bebían excesivamente. La compañía consideró que los costos del seguro se veían afectados por la corta vida de los bebedores excesivos.
La edad promedio al morir para 100 alcohólicos fue de 63.7 años con $s = 17.7$, mientras que 100 bebedores moderados y abstemios vivieron un promedio de 75.2 años, con $s = 8.7$. ¿Cómo interpretaría usted un intervalo de confianza del 95% para la diferencia en la vida promedio de los dos grupos?

47. Se realizó un experimento en la fijación de precios por parte de una cadena nacional de equipos estereofónicos. Durante un fin de semana, el precio de sus mejores reproductores de discos compactos se incrementó en un 4% en 35 tiendas y se redujo en un monto similar en otras 35 seleccionadas aleatoriamente. Los cambios en los ingresos por ventas se notaron en cada caso. En las tiendas que incrementaron su precio, los ingresos por ventas de reproductores de CD incrementaron en promedio US\$842, con $s =$ US\$217. El incremento promedio en los ingresos en aquellas tiendas que bajaron los precios fue de US\$817 con $s =$ US\$202. El gerente de marketing de la firma siempre consideró que un incremento en el precio aumentaría los ingresos más que una reducción en los precios (un concepto que los economistas llaman elasticidad de la demanda). ¿Qué dice al gerente un intervalo del 99% sobre los incrementos promedio en el ingreso?

48. Una teoría controvertida en los títulos y valores transados en las bolsas organizadas, siempre incrementan más los viernes que los lunes, debido al tiempo que toman las subastas de tesorería. Como proyecto prioritario, un especialista en finanzas en una gran universidad, selecciona aleatoriamente 302 acciones que se transan el viernes en la Bolsa de Valores de Nueva York y encuentra un cambio en el precio promedio de 0.375 puntos, con una desviación estándar de 0.075. Las 412 acciones seleccionadas aleatoriamente en las negociaciones del lunes produjo un cambio en el precio promedio de – 0.25 puntos, con una desviación estándar de 0.05. ¿Cómo cree usted que el especialista en finanzas puede interpretar los resultados de un intervalo del 99%?

49. Usted se acaba de graduar de la universidad y ha sido contratado como analista en control de calidad de Electric Charlie's, un gran productor de equipos de iluminación. Actualmente Electric Charlie's utiliza dos métodos para producir su sistema de iluminación casero Bright-Spot. Para determinar si un método es mejor que otros, usted selecciona 50 sistemas de cada método de producción. Los del primer método continúan proporcionando luz durante un tiempo promedio de 45.5 horas, con $s = 12.4$ horas. Los producidos utilizando el segundo método se fundieron a un promedio de 41.2 horas, con $s = 15.3$:

a. Calcule e interprete el intervalo de confianza del 95% para la diferencia entre los tiempos de servicio promedio de la población. Su supervisor desea que usted haga alguna recomendación. ¿Cuál sería su recomendación?

b. Sus muestras son lo suficientemente grandes como para estar 90% seguro de que el error no excede de dos horas? ¿De qué tamaño debería ser la muestra?

50. Como nuevo analista en la división de análisis financiero de una empresa con sede en Florida que fabrica jet skis, usted debe determinar si la empresa debería concentrar sus esfuerzos en el suministro a los clientes de la Costa Oeste o los de la Florida. La decisión dependerá en parte de cuál mercado paga el precio más alto. El CEO considera que el precio promedio en la Costa Oeste es más de US\$15 por encima de lo que la empresa puede recibir de los clientes de Florida. Utilizando estos datos, interprete para el CEO los resultados de un intervalo del 95%.

	Pedidos de la Costa Oeste	Pedidos de Florida
Número de pedidos	37	41
Precio promedio	US$418.10	US$397.20
Desviación estándar	73.00	62.10

51. Siete naciones asiáticas reportaron un incremento promedio en el ingreso *per cápita* del año anterior de US$121.20, con una desviación estándar de US$23.30. Nueve países europeos informaron cifras correspondientes de US$111.10 y US$19.10. Calcule e interprete el intervalo de confianza del 90% para la diferencia entre los incrementos promedio de la población en cuanto a ingresos se refiere. ¿Puede concluir que un continente parece haber generado incrementos más elevados que otro? No parece que las desviaciones estándar sean iguales.

52. A seis economistas que trabajan para el gobierno se les pide pronosticar las tasa de inflación para el año venidero. Ocho economistas que trabajan en empresas privadas reciben la misma tarea. Los seis economistas del gobierno reportan tasas del 4.2%, 5.1%, 3.9%, 4.7%, 4.9%, y 5.8%. Los ocho economistas privados proyectan tasas de 5.7%, 6.1%, 5.2%, 4.9%, 4.6%, 4.5%, 5.2% y 5.5%. ¿Cuál es su estimado de la diferencia en los pronósticos promedio de los dos grupos de economistas? Sea α 10% y se asumen varianzas iguales.

53. Muchos estudios sobre impacto económico se han realizado para determinar el efecto de los sindicatos laborales en las tasas salariales. Para manejar este importante asunto, un economista analiza 10 talleres agremiados en donde se encontró una tasa salarial promedio de US$22.07 y s = US$8.12. Doce talleres no agremiados revelan una media de US$24.17 y s = US$9.07. Utilice un intervalo del 99% para estimar si hay alguna diferencia en los niveles salariales promedio. ¿Debería sindicalizarse su taller? Se asumen varianzas iguales.

54. *The Wall Street Journal* publicó que un distribuidor de alimentos en el centro de Estados Unidos analizó los efectos de dos programas de ventas per cápita en el consumo de leche. Se dijo que diez ciudades tenían una publicidad extensiva en televisión, y se registró un incremento subsiguiente en el consumo promedio diario de 0.25 galones y una desviación estándar de 0.09 galones. Otras doce ciudades fueron saturadas con anuncios en los periódicos. Allí resultó un incremento de 0.02 galones en el consumo promedio per cápita, con $s = 0.02$ galones. Si se asume que las varianzas son iguales, ¿cómo interpretaría usted los resultados de un intervalo del 90%?

55. Como director de producción de Maxx Manufacturing, usted debe decidir cuál de las dos plantas debe responsabilizarse de producir los corchos para el vino utilizados en Paul Masson Wineries. Esta decisión se fundamentará en los niveles de productividad. Una muestra de 67 días en la planta de Northridge produjo una media de 92.2 miles de corchos por día con $s = 12.2$ millones. La planta de Southridge produjo un promedio de 89.2 millones de corchos con $s = 15.4$ durante 54 días:

a. Pruebe la hipótesis de igualdad de las medias a un nivel de significancia del 10%.

b. ¿Qué tamaño deben tener las muestras para estar 90% seguros de que el error no excede de 5,000 corchos?

56. Muchas decisiones financieras de las empresas se basan en los flujos de caja. Una máquina antigua que su empresa está utilizando actualmente, generó un flujo de caja neto positivo promedio de US$15.600 con una desviación estándar de US$2.300 para un período de 18 semanas. Una nueva máquina utilizada por la competencia proporcionó los valores respectivos de US$12.100 y US$3.400 durante un período de 13 semanas. Pruebe la hipótesis de igualdad de las medias a un nivel de significancia del 2%. Con base en este análisis, ¿qué recomendaría usted? Se asume $\sigma_1^2 \neq \sigma_2^2$.

57. Muchos países europeos utilizan un impuesto al valor agregado (IVA), el cual es un impuesto sobre el valor agregado a un bien en cada etapa de su producción. Ocho países que utilizan un IVA sobre el consumo reportaron un ingreso promedio semanal per cápita de US$1,142, con s = US$312. Diez países que utilizan un IVA sobre el ingreso bruto, reportaron un impuesto semanal per cápita promedio de US$1,372, con s = US$502. Si $\alpha = 0.05$ y $\sigma_1 \neq \sigma_2$, ¿cómo interpretaría los resultados de una hipótesis $\mu_1 = \mu_2$?

58. El impacto de los diferentes métodos de remuneración sobre la productividad y los niveles de satisfacción de los trabajadores, siempre ha sido de interés para los economistas laborales. *Fortune* publicó que una empresa de productos deportivos experimentó los efectos de dos métodos de remuneración sobre la moral de los empleados, en una planta de Ohio. Catorce trabajadores a quienes se les pagó un salario fijo presentaron una prueba que medía la moral y obtuvieron un puntaje promedio de 79.7 con $s = 8.2$ Doce trabajadores a quienes se pagaba sobre comisión lograron una media de 72.7 con $s = 5.1$. Sea $\alpha = 0.10$ y se asume $\sigma_1 \neq \sigma_2$, ¿qué puede concluirse respecto a los méritos relativos de los dos sistemas de pago con base en la prueba de hipótesis resultante?

59. El director financiero (CFO) de una de las firmas que aparece en la lista de las mejores 500 empresas de la revista *Fortune*, debe decidir qué forma de financiación es menos costosa: si la financiación de la deuda o el financiamiento por venta de participación. Ella analiza las transacciones recientes del mercado de firmas similares a la de ella y descubre que 17 empresas que utilizan bonos (financiación de la deuda) experimentaron un costo promedio del 17.3% con $s = 3.7\%$, y 10 emisiones de acciones recientes (financiamiento por venta de participación) terminó en cifras de 22.7% y 4.7%, respectivamente. Puede usted ayudarle a probar la hipótesis apropiada a un nivel de significancia del 5%, si $\sigma_1^2 \neq \sigma_2^2$. ¿Proporciona este estudio alguna evidencia sobre cuál es el método de financiación menos costoso?

60. B.F. Skinner, notable teórico comportamentalista, defendió el uso del refuerzo positivo para moldear las actitudes en el trabajo. Texaco, Inc., ha utilizado las técnicas de Skinner. Una muestra de 45 empleados que recibieron este tratamiento positivo obtuvo un puntaje promedio de 5.5 sobre 10 en una prueba de actitudes, con $s = 1.2$. Sesenta empleados que trabajan en un área que no recibe refuerzo positivo obtuvo un promedio de 4.8 en la prueba, con $s = 1.9$:

a. Calcule e interprete el intervalo de confianza del 98% para la diferencia entre las medias poblacionales. ¿Los resultados parecen apoyar las teorías de Skinner?

b. ¿Cuáles son los tamaños apropiados para las muestras si usted desea estar 95% seguro de que el error es menos que 1 punto?

61. Un contratista de plomería desea estimar la diferencia en la proporción de trabajos de construcción que requieren una segunda visita a la obra de construcción antes de finalizar el trabajo. El contratista intenta entregar su trabajo a la empresa de plomería con el nivel más bajo de visitas repetidas. De los 50 trabajos terminados por Alpha Plumbing Corporation, 39 requirieron una segunda visita, mientras que Omega Plumbing Group hizo una segunda visita en el 67% de sus 60 trabajos. ¿Qué puede usted concluir sobre la diferencia en las proporciones de trabajos que requieren un segundo esfuerzo, con base en un intervalo del 95%? ¿Cuál compañía debería utilizar el contratista, o esto hace alguna diferencia?

62. Dos conjuntos "idénticos" de 50 empleados se inscriben en dos programas diferentes de capacitación y luego se les practica una prueba de aptitudes. La diferencia promedio en los puntajes es de 13.5, con una desviación estándar en tales diferencias de 4.3. ¿Qué se concluye con un intervalo del 95% sobre la relativa eficacia de los programas de formación adicional?

63. Blanca Nieves compra a sus siete enanos nuevas palas para navidad. La cantidad que cada enano puede excavar en la mina con las palas viejas y las palas nuevas aparece a continuación. Pruebe la hipótesis apropiada a un nivel del 10%. ¿El regalo de Blanca Nieves para sus siete pequeños amiguitos mejorará la producción?

	Producción diaria en toneladas	
Enano	**Palas vieja**	**Palas nuevas**
Doc	1.7	1.9
Feliz	1.4	1.5
Gruñón	2.1	2.2
Tímido	1.9	2.0
Dormilón	2.2	2.2
Tontín	1.4	1.5
Estornudo	1.9	1.8

64. En finanzas, un mercado eficiente se define como el que adjudica fondos para el uso más productivo. Existe una bibliografía considerable que está diseñada para determinar si los mercados de valores todavía son eficientes. *Business Week* recientemente aplicó encuestas a los analistas financieros. De 110 analistas que trabajaban para empresas productoras privadas en su esfuerzo por vender los valores de sus firmas, 42 consideraron que los mercados eran eficientes, mientras que 31 de los 75 analistas que trabajaban para casas de corretaje que asisten tales ventas aceptaron que los mercados eran eficientes. Pruebe la hipótesis a un nivel del 5%; ¿hay alguna diferencia en la proporción de estos dos tipos de analistas que aceptan el concepto de eficiencia del mercado?

65. Se va a probar dos medicinas para establecer las reacciones adversas en los pacientes. Se suministra Accura a 37 pacientes y se encuentra que el 25% tienen una reacción. De las 55 personas que recibieron Tardi, el 29% experimentó una reacción. Sea α 1%. ¿Qué puede usted concluir del intervalo resultante?

66. Muchas empresas grandes utilizan los servicios de empresas consultoras en el proceso de selección de empleados. Para probar el beneficio de tales consultores, la IBM comparó recientemente 100 empleados contratados a través de su propia división de selección de personal, con 150 empleados contratados de manera menos formal. Los resultados mostraron que el 55% de los del primer grupo no avanzaron más allá del nivel administrativo intermedio en siete años de trabajo, mientras que la cifra correspondiente para el segundo grupo fue del 60%. ¿Qué puede concluir sobre la diferencia en la efectividad de estos dos métodos de contratación con base en un intervalo del 98%?

67. Cada unidad de negocio estratégico (SBU) dentro de una compañía es responsable de desarrollar su propia estrategia. En dicho esfuerzo, Whirlpool Corporation fue una de las primeras compañías norteamericanas en enfatizar la estrategia del marketing estratégico. Un método común fue el de una encuesta al consumidor. Si Whirlpool encontró que el 28% de los 70 hombres a quienes se les practicó la encuesta dieron su aprobación para electrodomésticos pintados, mientras que el 34% de las 80 mujeres lo aprobaron. ¿Qué revela un intervalo del 90% sobre la diferencia entre la proporción de hombres y mujeres que prefieren las lavadoras y secadoras pintadas?

68. De 35 personas que están en un plan de adelgazamiento, el 70% alcanzó la meta. Un segundo plan funciona para el 65% de las 50 personas que lo utilizan:

a. ¿Un intervalo del 99% indica una diferencia significativa en la tasa de éxito de los planes?

b. ¿Qué tamaño han de tener las muestras para estar seguro con un nivel de confianza del 99% de que el error no excederá del 5%?

69. United Airlines descubre que la mitad de una muestra de 150 vuelos llegan a tiempo cuando vuelan del este al oeste. De 160 vuelos que van en dirección este, 72 llegan a tiempo. Sea α 10%.

a. ¿Qué dice el intervalo resultante sobre la probabilidad relativa de llegar a tiempo dependiendo de la dirección del vuelo?

b. Su jefe, que en esta semana vuela por United Airlines, desea estar 90% seguro de que el error en la estimación no exceda del 10%. ¿Qué tan grandes deberían ser las muestras?

70. Como parte del proyecto prioritario, un especialista en mercadeo de la Universidad de North Texas State en Denton, Texas, encuestó en un centro comercial local a 100 hombres y 100 mujeres respecto a sus hábitos en las compras. De los hombres, 79 dijeron que habían utilizando una tarjeta de crédito para hacer compras superiores a US$10 en el último mes, mientras que 84 mujeres admitieron hacer este mismo tipo de pago. El estudiante estaba intentando refutar el hecho de que es más probable que las mujeres utilicen la tarjeta de crédito. A un nivel del 5%, ¿lo logró? Plantee y pruebe la hipótesis apropiada.

71. Un ensayo realizado por un especialista en ciencias de la computación en Ohio State University, titulado "Su chip está por llegar", analizó la calidad de los chips de computador fabricados por dos compañías. De 453 chips producidos por la compañía 1, 54 fueron defectuosos. De 317 producidos por la compañía 2, 43 fueron defectuosos. Si $\alpha = 0.10$, ¿existe evidencia que sugiera que una empresa mantiene un control de calidad más estricto que la otra, con base en la prueba de la hipótesis apropiada?

72. Denny Dimwit, un analista de valores de Your Bottom Dollar, Inc. siempre ha considerado que es más probable que los bonos convertibles se sobrevaloren más que los bonos de ingresos. De 312 bonos convertibles analizados el año anterior, Denny encontró que 202 estaban sobrevalorados, mientras que 102 de los 205 bonos de ingresos estaban sobrevalorados. ¿Estos datos confirman la suposición de Denny? Sea $\alpha = 0.10$, y pruebe la hipótesis.

73. Se utilizan dos métodos de producción para ensamblar reproductores de discos compactos. El tiempo promedio requerido por ambos métodos se ha estimado en 5.6 minutos aproximadamente. Sin embargo, los estudios parecen sugerir que las varianzas en tales tiempos difieren. Es importante mantener horarios de producción similares para coordinar los itinerarios de producción. Determine si las varianzas en los tiempos de producción no son similares. Quince reproductores de discos compactos producidos por el primer método reportan una desviación estándar de 5.4 minutos, y 17 reproductores del segundo método reportan una desviación estándar de 4.8 minutos. A un nivel de significancia del 5%, qué concluiría usted de la prueba de la hipótesis apropiada?

74. Se recolectan datos para determinar si hay diferencia en las varianzas de los ingresos diarios en las dos tiendas. Dados los siguientes datos, ¿cuál es su conclusión con base en una prueba de hipótesis siendo α 1%?

Tienda 1	Tienda 2	Tienda 1	Tienda 2
US$45.78	US$67.89	US$12.55	US$34.91
34.66	76.45	37.77	56.88
65.89	87.12	21.87	45.99
54.78	98.65	23.45	
98.66	65.87	56.98	

PUESTA EN ESCENA

El escenario de apertura en la sección *Escenario* describió su trabajo como analista internacional: usted debe comparar las oportunidades de inversión en Europa y Asia. Usted debe preparar un informe para ser utilizado por los directivos de la empresa para decidir en cuál área deben hacer la mayoría de inversiones; la decisión dependerá de varios factores. Su compañía desea concentrarse en donde las condiciones promedio sean más favorables con respecto a 1) tasas de rendimiento, 2) niveles de utilidad, 3) niveles de inversión, 4) tasas de crecimiento económico, y 5) medidas de la estabilidad política. Las tasas promedio de incumplimiento en los créditos e inversiones también se deben considerar. Las estimaciones de las diferencias entre estas medias también son críticas.

Los datos que aparecen a continuación están en un archivo llamado ABROAD en su disco de datos. Fueron recolectados por su sección Análisis de datos para los datos trimestrales durante los últimos 20 trimestres. RETE y RETA son tasas de rendimiento en Europa y Asia respectivamente, medidas en porcentajes. De igual forma, INVE e INVA son inversiones en billones de dólares, PROE y PROA son utilidades en billones de dólares, GROE y GROA son tasas de crecimiento económico en porcentajes, LICE y LICA son evaluaciones de estabilidad política medida de 1 a 10 en la escala de Likert suministrada por 20 analistas políticos. A cada analista se le pidió clasificar tanto Europa como Asia respecto a la estabilidad política. Entre menor sea la clasificación, se determinará que los gobiernos extranjeros son menos estables.

De 25 inversiones recientes en Europa, dos experimentaron fallas que resultaron en una pérdida total de todo el dinero invertido. En Asia, 27 proyectos de inversión presentaron siete fallas.

Proporcione todo el análisis estadístico necesario y una discusión completa de la interpretación y de los resultados. Establezca claramente sus conclusiones y recomendaciones respecto al área, Europa o Asia, en la cual su firma debe concentrar su actividad inversionista.

RETE	RETA	INVE	INVA	PROE	PROA	LICE	LICA	GROE	GROA
20	15	250	52	10	3	5	2	9	9
25	17	240	54	14	6	8	4	8	8
27	14	210	74	15	5	7	2	9	5
26	12	195	84	17	2	5	3	12	7
12	8	174	147	21	5	6	6	15	5
15	9	154	54	25	4	9	2	14	6
16	7	214	95	21	7	8	1	17	5
24	5	187	87	26	8	5	5	14	8
26	14	165	85	18	9	4	2	12	5
29	14	240	96	17	5	7	3	11	9
31	15	287	74	15	4	8	2	15	11
14	18	247	35	19	7	5	5	12	12
15	5	265	68	14	8	6	4	14	8
14	8	187	54	15	5	8	1	16	9
18	7	198	78	18	11	5	2	10	6
9	12	177	104	12	10	6	5	11	8
8	14	154	108	21	14	8	6	12	7
15	10	147	147	11	15	4	1	14	11
16	11	165	158	10	9	7	4	11	5
24	11	147	211	9	8	8	5	11	4

Del escenario a la vida real

Una visita al U.S. Census Bureau (*www.census.gov*) le ayudará a familiarizarse con el comercio exterior como componente de la economía de Estados Unidos. De Census Bureau Home Page, seleccione "Subjects from A to Z" (Temas de la A la Z). Allí, vaya por el índice hasta "foreign trade" (Comercio exterior) y haga clic. Seleccione "Trade Balances with U.S. Trading Partners" (Balanzas comerciales con países que mantienen relaciones comerciales con Estados Unidos) bajo la lista de características. Luego, bajo el encabezado Trade Balances with Partner Countries— Current Year (Balanzas comerciales con países socios – Año actual), haga clic en "Try it out" (Prueba) (Una versión avanzada de Netscape automáticamente saltará la página con el paso "prueba"). Luego vaya al área "Select an item" (Seleccionar un ítem) para el cual usted desea la balanza comercial.

Usted podrá seleccionar de una lista un país en particular y ver sus datos mensuales sobre exportaciones, importaciones y balanzas comerciales con los Estados Unidos. Datos de doce meses y cifras anuales suministradas. Seleccione dos países africanos, dos asiáticos y dos europeos. Analice los datos mensuales para cada país. ¿Las balanzas comerciales cambian frecuentemente de positivo a negativo? Anote sus observaciones y la balanza comercial anual de cada país. ¿Qué tan similares son las balanzas comerciales de Estados Unidos y las de los dos países africanos? ¿Los dos países asiáticos? ¿Y los dos países europeos? Compare las balanzas africanas, asiáticas y europeas tomando a la vez dos países de diferentes áreas en cada comparación. Usted puede detectar cualquier patrón bien sea por tamaño o signo de las balanzas anuales?

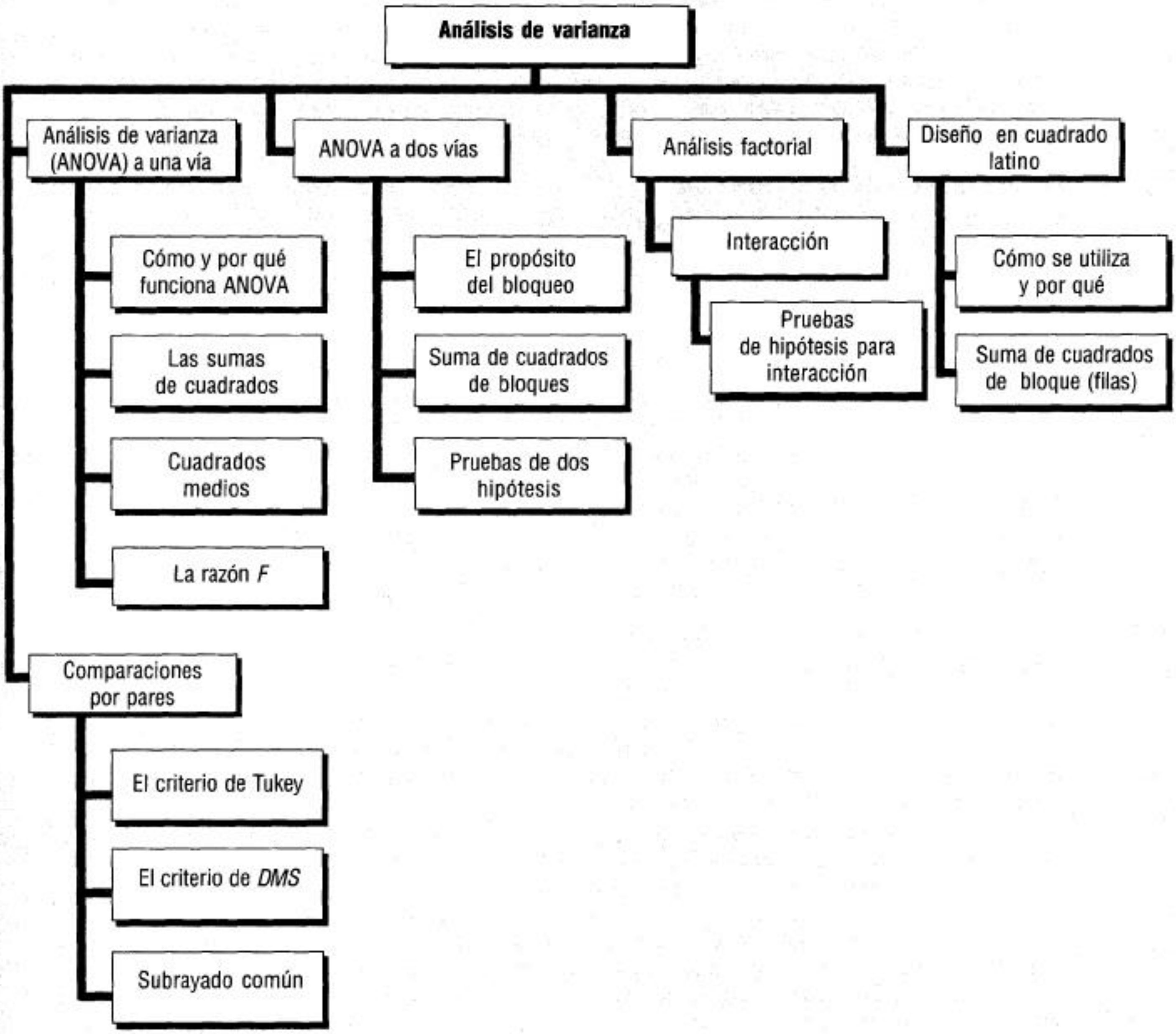
En esta misma área del sitio web también se encuentra disponible información detallada sobre los productos comercializados con cada país. Mire el encabezado, Country Detail (Detalle por País), y luego haga clic en el año más reciente. Allí, observe y haga clic en cada uno de los seis países de los cuales usted obtuvo la información sobre la balanza comercial. ¿Hay diferencias notables entre los países respecto a los productos que comercian con Estados Unidos?

10

Análisis de varianza

Plan del capítulo

El análisis de varianza se utiliza para probar hipótesis sobre la igualdad de tres o más medias poblacionales. Al comparar las varianzas muestrales, es posible sacar alguna conclusión o inferencia sobre los valores relativos de las medias poblacionales.



ESCENARIO

En la emisión del mes de junio de 1997, el *U.S. News and World Report* publicó un informe de la Agencia Central de Inteligencia (CIA), detallando el desempeño económico de las economías más grandes del mundo durante1995. El Grupo de los Siete (G-7), denominado Cumbre de los Ocho desde la inclusión de Rusia, se reunió en Denver en 1997 para discutir las formas de combatir la pobreza mundial. El interés se concentró en el cambio del estado de las economías mundiales y el establecimiento de políticas económicas y políticas que promovieran el desarrollo global.

La siguiente tabla, compilada por la CIA, antes de la Cumbre del Grupo de los Siete, proporciona una lista de las 10 economías más grandes del mundo con el producto interno bruto real (PIB).

Como se han producido varios cambios de posición entre las naciones durante los últimos años, la discusión en Denver se centró en los giros del orden de la economía mundial. "Se planteó una pregunta respecto a si había alguna diferencia significativa descendente en los tamaños de las economías". Los líderes de las naciones del grupo de los siete consideraron que los niveles de inflación y desempleo enumeradas en la sección "Puesta en escena" que se encuentra al final de este capítulo, fueron de especial importancia al medir el bienestar económico de una nación. El material presentado en este capítulo será de gran utilidad al tratar estos asuntos.

Rango	País	PIB (miles de millones de dólares de Estados Unidos)	Rango	País	PIB (miles de millones de dólares de Estados Unidos)
1	Estados Unidos	$7,248	6	Francia	$ 1,173
2	China	3,500	7	Inglaterra	1,138
3	Japón	2,679	8	Italia	1,089
4	Alemania	1,452	9	Brasil	977
5	India	1,409	10	Rusia	796

10.1 Introducción

En el capítulo 9 se contrastaron las hipótesis respecto a la igualdad de dos medias poblacionales. Infortunadamente, estas pruebas estaban restringidas en su aplicación a una comparación de sólo dos poblaciones. Sin embargo, muchas decisiones en los negocios requieren de la comparación de más de dos poblaciones. Es aquí donde el análisis de varianza (ANOVA) es de gran utilidad.

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.

Un estudio completo de modelos de efectos aleatorios va más allá del alcance de este texto. La intención de este capítulo se concentrará en los modelos de efectos fijos.

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

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

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

$$= 82.14$$

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,

$$\text{Variación total} = \text{variación del tratamiento} + \text{variación del error}$$

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
$$SCTR = \Sigma r_j(\overline{X}_j - \overline{\overline{X}})^2 \qquad [10.4]$$

El número de observaciones o filas en cada tratamiento, r_j, se multiplica por las diferencias cuadradas entre la media de cada tratamiento, $\overline{X}_j$, 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

$$\begin{aligned} SCE = {} & (85 - 80)^2 + (72 - 80)^2 + (83 - 80)^2 + (80 - 80)^2 \\ & \text{Para el primer tratamiento} \\ & + (80 - 81)^2 + (84 - 81)^2 + (81 - 81)^2 + (78 - 81)^2 + (82 - 81)^2 \\ & \text{Para el segundo tratamiento} \\ & + (82 - 85)^2 + (80 - 85)^2 + (85 - 85)^2 + (90 - 85)^2 + (88 - 85)^2 \\ & \text{Para el tercer tratamiento} \\ = {} & 186.0 \end{aligned}$$

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:

$$\begin{aligned} CMT &= \frac{SCT}{n-1} \\ &= \frac{251.7}{14-1} \\ &= 19.4 \\ CMTR &= \frac{SCTR}{c-1} \\ &= \frac{65.7}{3-1} \\ &= 32.9 \\ CME &= \frac{SCE}{n-c} \\ &= \frac{186.0}{14-3} \\ &= 16.9 \end{aligned}$$

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	$F = \dfrac{CMTR}{CME}$	[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 $\mu_1 = \mu_2 = \mu_3$.

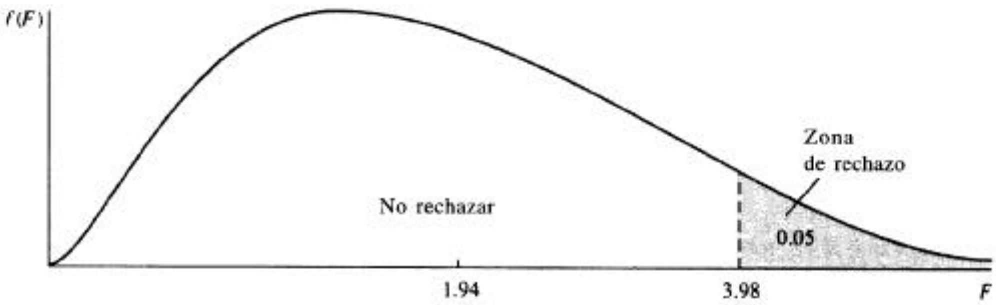
El valor crítico de F que es considerado significativamente grande puede encontrarse en la tabla G (apéndice III) igual que antes. Se asume que el CEO desea probar las siguientes hipótesis a un nivel del 5%:

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

H_A: 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

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

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

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

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

$$\begin{aligned}
CMTR &= \frac{55.33}{3} \\
&= 18.44 \\
CME &= \frac{5.67}{24} \\
&= 0.236
\end{aligned}$$

Entonces la razón F es

$$\begin{aligned}
F &= \frac{CMTR}{CME} \\
&= \frac{18.44}{0.236} \\
&= 78.14
\end{aligned}$$

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		

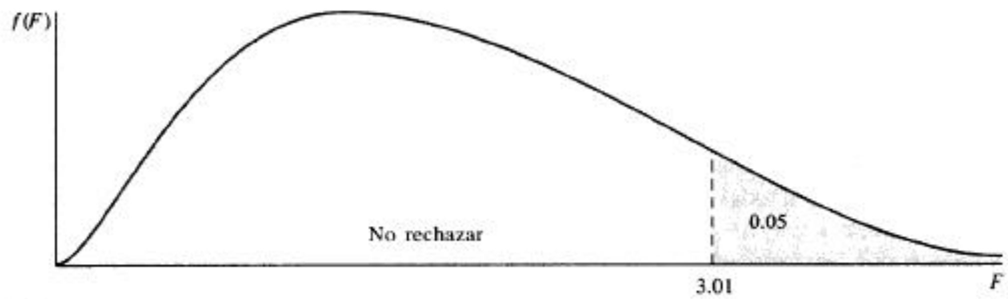
H_0: $\mu 1 = \mu 2 = \mu 3 = \mu 4$

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

$$\begin{aligned} T &= 3.90\sqrt{\frac{0.236}{7}} \\ &= 0.716 \end{aligned}$$

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.

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

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

$$DMS = \sqrt{\frac{2(0.236)4.26}{7}}$$
$$= 0.536$$

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: $\mu1 = \mu2 = \mu3$

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

$$|\bar{X}_c - \bar{X}_f| = |35.17 - 30.25| = 4.92 < 5.85$$
$$|\bar{X}_c - \bar{X}_b| = |35.17 - 24.20| = 10.97 > 5.48$$
$$|\bar{X}_f - \bar{X}_b| = |30.25 - 24.20| = 6.05 < 6.08$$

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.

$\bar{X}_b$	$\bar{X}_f$	$\bar{X}_c$
24.2	30.25	35.17

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.

Ejercicios de la sección

1. Un productor de pinturas para vivienda desea comparar el factor brillo de su pintura utilizando cuatro emulsiones diferentes. Se pintan cinco tablas con cada tipo de emulsión y la clasificación que se dio a cada una aparece aquí.

	Emulsión			
Tablas	1	2	3	4
1	79	69	83	75
2	82	52	79	78
3	57	62	85	78
4	79	61	78	73
5	83	60	75	71

   a. ¿A un nivel del 1%, existe alguna diferencia en la clasificación promedio?

   b. Use el método de Tukey para probar las diferencias y determinar si el productor debería utilizar algún tipo o evitar su uso?

2. Un estudio reciente realizado por American Assembly of Collegiate Schools of Business comparó los salarios iniciales de los nuevos graduados en diversos campos. Una parte de sus resultados se representa en la tabla siguiente. A un nivel del 5%, ¿parece que hay diferencia en los salarios promedio (en miles de dólares) de los graduados en los diferentes campos? (sistemas de información por computador–SIC, y métodos cuantitativos–MC.)

	Campo de estudio			
Graduado	Finanzas	Mercadeo	SIC	MC
1	23.2	22.1	23.3	22.2
2	24.7	19.2	22.1	22.1
3	24.2	21.3	23.4	23.2
4	22.9	19.8	24.2	21.7
5	25.2	17.2	23.1	20.2
6	23.7	18.3	22.7	22.7
7	24.2	17.2	22.8	21.8

3. Considerando sus resultados del problema anterior, utilice el método de Tukey para determinar cuáles medias son diferentes. ¿Obtiene los mismos resultados con el método *DMS*? Mantenga $\alpha = 0.05$. Resuma los resultados con el subrayado común.

4. Una empresa de suministros médicos desea comparar la producción diaria promedio de sus tres plantas en Toledo, Ohio; Ottumwa, Iowa; y Crab Apple Cove, Maine. Se recolectaron datos para cada sitio y aparecen

a continuación. ¿A un nivel del 10%, existe diferencia en las medias? Las cifras están dadas en unidades de producción.

Toledo: 10, 12, 15, 18, 9, 17, 15, 12, 18

Ottumwa: 15, 17, 18, 12, 13, 11, 12, 11, 12

Crab Apple Crove: 12, 17, 15, 15, 18, 12, 13, 14, 14

5. Ralph trabaja como mesero para estudiar taxidermia en la universidad. Las propinas que recibió recientemente en tres restaurantes aparecen a continuación:

Beef & Boar	Sloppy Sam's	Crazy Charlie's
US$5.12	US$5.60	US$6.00
6.18	4.25	5.00
4.25	6.25	2.00
5.00	7.25	4.50
6.00	5.00	6.50
3.25	4.00	5.50

¿En cuál sitio debería trabajar si desea maximizar sus propinas, o parece no haber diferencia? Sea $\alpha = 5\%$.

6. Debido a que los corredores de bolsa trabajan en su mayoría por comisión, están interesados en la actividad comercial del mercado. Se realizó un estudio para determinar si había una diferencia en las comisiones promedios pagadas con base en el día de la semana, y se recolectaron los datos que se muestran aquí, en cientos de dólares. A un nivel del 1%, ¿qué días parecen pagar más? Utilice el criterio de Tukey. Resuma con el subrayado común.

Lunes	Martes	Miércoles	Jueves	Viernes
US$21	28	11	15	25
26	21	14	14	23
24	19	12	12	26
32	15	10	12	28
25	12	10	16	24
26	10	12	13	25
24	13	15	18	29

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
$\bar{X}_j$	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

$$SCBL = \Sigma c_i(\bar{X}_i - \bar{\bar{X}})^2 \qquad [10.13]$$

El número de tratamientos en cada bloque , c_i, se multiplica por la diferencia al cuadrado entre la media para cada bloque, $\overline{X}_i$, 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,

$$\begin{aligned} \text{Cuadrado medio total} = CMT &= \frac{SCT}{n-1} \\ &= \frac{806.93}{14} \\ &= 57.64 \\ \text{Cuadrado medio del tratamiento} = CMTR &= \frac{SCTR}{c-1} \\ &= \frac{0.93}{2} \\ &= 0.47 \end{aligned}$$

En ANOVA a dos vías,

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

$$= \frac{40.96}{8}$$
$$= 5.1$$

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

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

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

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

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:

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

$$H_A\text{: No todas las medias de las filas son iguales}$$

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

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

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)

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

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:

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

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 >
```

Ejercicios de la sección

7. Como recién graduado con un título en mercadeo, usted acaba de ser aceptado en un gran trabajo con una importante empresa de cosméticos de New York. Debe ayudar en el análisis de la efectividad de tres presentaciones publicitarias. Se seleccionan aleatoariamente cinco clientes. A cada uno se le presenta un anuncio y se le pide calificarlo. A continuación se muestran los resultados. Su supervisor no está seguro de cómo proceder, pero sus amplios conocimientos en análisis estadístico, con base en su trabajo en el curso de la universidad, le dice que el análisis de varianza a dos vías es apropiado ya que cada cliente (bloque) calificó los tres anuncios (tratamientos). ¿Existe alguna diferencia significativa en las actitudes de los consumidores? Sea $\alpha = 0.01$.

	Presentación publicitaria		
Consumidor	**1**	**2**	**3**
1	50	45	45
2	45	30	35
3	30	25	20
4	45	35	40
5	40	30	35

8. Debits and Credits Inc., una firma de contabilidad en Rocky Top, Tennessee, tiene la política de evaluar a cada nuevo empleado haciendo que complete varias afirmaciones contables y que compile todo error. Usted y otros dos miembros nuevos de las empresas, Seymore Nueshaum y Gretchen Nordick, deben llenar las seis declaraciones. Esto lo revisa un socio antiguo de la empresa y los errores que cada uno de ustedes cometió aparecen en la siguiente tabla. Parece que uno de ustedes puede ser más o menos eficiente en sus destrezas contables. Sea $\alpha = 0.05$. Si es así, ¿cuál es? Se decide hacer bloques en cada afirmación para dar cuenta de toda diferencia en la dificultad que pueda existir.

	Número de errores		
Afirmación	**Usted**	**Seymore**	**Gretchen**
1	2	2	3
2	1	3	4
3	0	1	4
4	4	6	5
5	2	3	4
6	1	4	3

9. Las negociaciones actuales entre el sindicato y la gerencia se concentran en el efecto que tienen los métodos de remuneración en la producción del trabajador. Una gran empresa tiene cinco plantas. En cada una, a los trabajadores se les paga por comisión, salario, o un plan de bonificaciones. Tres trabajadores, seleccionados aleatoriamente de todas las plantas, recibieron cada uno un método diferente de pago. Su producción diaria medida en unidades aparece aquí. Se piensa que es necesario bloquear las plantas corrigiendo toda diferencia que pueda existir en la producción promedio de la planta.

Con base en estos datos, ¿cuál plan de pagos sugeriría a la gerencia si el objetivo es maximizar la producción? Sea $\alpha = 0.05$.

	Método de pago		
Planta	**Comisión**	**Salario**	**Bonificaciones**
1	25	25	37
2	35	25	50
3	20	22	30
4	30	20	40
5	25	25	35

10. Una empresa carbonífera en West Virginia planea analizar la producción promedio de tres minas. Cuatro grupos de trabajo trabajarán en cada mina y registrarán en toneladas la producción de carbón resultante. Debido a que cada grupo trabajará en cada mina, se utilizará el análisis de varianza a dos vías haciendo bloqueo en los grupos. Como nuevo supervisor administrativo, usted debe determinar si existe alguna diferencia en la productividad de las minas. Sea $\alpha = 0.01$. Señale cuáles minas son más productivas con un subrayado común.

Grupo	Mina 1	Mina 2	Mina 3
1	42.7	54.1	56.9
2	47.1	59.2	59.2
3	32.1	53.1	58.7
4	29.2	41.1	49.2

11. Speedo fabrica bicicletas de turismo para verdaderos ciclistas. El ingeniero jefe de control de calidad decide comparar las velocidades más altas obtenidas, utilizando tres mecanismos de cambios diferentes; 5 ciclistas experimentados son cronometrados cuando corren con cada uno de los tres mecanismos, los resultados aparecen aquí. ¿Los datos sugieren una diferencia en las velocidades promedio a un nivel del 1%?

	Mecanismo		
Ciclista	1	2	3
1	40	51	37
2	42	49	38
3	37	53	38
4	45	57	41
5	42	42	40

12. The U-Plant'um Nursery debe determinar si existe alguna diferencia en la tasa de crecimiento de los árboles jóvenes que se han tratado con fórmulas de químicos diferentes. Debido a que la condición del suelo se debe considerar, los árboles jóvenes tratados con cada fórmula se plantan en cada uno de tres tipos de suelo. Las tasas de crecimiento resultantes durante un período determinado aparecen a continuación. ¿Parece existir una diferencia en el factor de crecimiento de las fórmulas después de hacerse la corrección sobre el suelo? Sea α 1%.

	Fórmula			
Tipo de suelo	1	2	3	4
Arenoso	10	8	5	7
Arcilloso	12	15	17	14
Rocoso	17	16	15	15

13. Curly, Moe, y Larry venden tenedores eléctricos puerta a puerta. Cada uno va independientemente a cuatro barrios y entregan sus propios registros de ventas. El número de tenedores eléctricos vendidos se encuentra registrado en la tabla. A un nivel del 5%, ¿parece que uno de los vendedores tiene un futuro más brillante que los otros? Si es así, ¿cuál es? Debido a que cada vendedor visitó todos los vecindarios, haga una prueba para ver si debería utilizarse el bloqueo.

Vecindario	Curly	Moe	Larry
1	15	12	19
2	27	25	12
3	24	29	30
4	32	31	29

14. El Instituto Nacional de Salud practicó encuestas a 1,060 adultos para determinar cómo pasaban su tiempo libre. Los datos han sido desglosados por grupos de edad y se han condensado a sólo 16 observaciones para efectos de computación. ¿Parece haber alguna diferencia en el tiempo promedio que se pasa en las diferentes actividades? Las observaciones están en horas por semana. Haga una prueba para determinar si debería utilizarse el bloqueo. Sea $\alpha = 0.05$

	Actividad			
Encuestados (por edad)	TV	Lectura	Deportes	Tiempo con la familia
15-18	35	12	10	6
19-25	22	13	12	8
26-35	25	15	8	15
36 y más años	27	20	5	20

10.5 Análisis factorial

En la discusión sobre el análisis de varianza a dos vías, se reconoció la presencia de un segundo factor que influenciaría las unidades experimentales. Ya que no se tuvo interés en analizar esta segunda fuerza, se buscó eliminar este impacto. En un ejemplo anterior sobre la productividad del sistema, se temía que la experiencia contaminara los resultados del estudio y por ello se bloqueó el nivel de experiencia.

Se asume ahora que se desea no solamente probar el sistema de computación, sino también comparar el efecto de los dos paquetes de software distintos en la producción. Por tanto, se desea probar simultáneamente el efecto de dos factores diferentes en la producción: el sistema de computación y el paquete de software. El diseño experimental apropiado que se empleará se denomina **análisis factorial**, o análisis de varianza de dos factores.

Análisis factorial Evaluación de dos factores de interés al mismo tiempo.

En análisis factorial se puede efectuar una prueba de **efectos principales** por cada factor. Cada prueba es muy parecida como las presentadas anteriormente: se diseñan para determinar si niveles diferentes de cualquier factor tienen impacto en las unidades experimentales en diversas formas. Si no se encuentran efectos principales para un factor, la hipótesis nula no se rechaza.

Prueba de efectos principales Pruebas sobre ambos factores para determinar si los niveles diferentes del factor influyen en las unidades de manera diversa.

Vale la pena notar que en análisis factorial existen dos aspectos a considerar. Cada factor tiene más de un nivel. Se puede decir que el factor *A* tiene *a* niveles y el factor B tiene *b* niveles. En este caso en particular se desea comparar tres sistemas de computadores y dos paquetes de software. Por consiguiente, el factor A tiene tres niveles y el factor B tiene dos niveles. Hay a X b, o seis combinaciones del sistema/paquete. Esto es lo que se denomina como diseño factorial 3 X 2. Cada combinación se llama *tratamiento*. Ahora hay seis tratamientos diferentes en consideración. Cada tratamiento, o combinación de sistema/paquete, se muestra en una de las seis celdas que aparecerán en la tabla.

En los estudios anteriores sobre análisis de varianza, cada celda contenía sólo una observación. Por ejemplo, sólo una persona con una cantidad de experiencia dada registró su producción en la celda apropiada. Sin embargo, para realizar el análisis factorial, *debe* aparecer más de una observación en cada celda. El número de observaciones dentro de una celda es lo que con frecuencia se denomina número de **replicaciones**, *r*. La tabla 10.5 lo ilustra. En cada una de las seis celdas aparecerá la producción de varias unidades experimentales (empleados). El primer empleado que trabaja en el sistema 1 y el paquete 1 (S1/P1) puede producir 27 unidades de producción. Si otros dos empleados se seleccionan para trabajar en S1/P1, pueden producir 26 y 25 unidades respectivamente. El promedio de esa celda, μ_{11}, es igual a 26. La prueba debe diseñarse de manera que cada celda tenga el mismo número de observaciones. Un número desigual va más allá del alcance de este texto. Toda la tabla aparece como la tabla 10.5, con tres observaciones en cada una de las seis celdas, resultan 18 unidades experimentales.

Tabla 10.5
Configuración de las celdas para el análisis factorial

		FACTOR A (sistema de computación) Tres niveles del factor A		
		1	2	3
FACTOR B (paquete de software)	1	27 26 25 $\mu_{11} = 26$	20 22 21 $\mu_{12} = 21$	30 26 28 $\mu_{13} = 28$
Dos niveles del factor B	2	28 27 29 $\mu_{21} = 28$	20 26 23 $\mu_{22} = 23$	30 31 29 $\mu_{23} = 30$

Se dice que el diseño es *completamente aleatorizado* si las 18 unidades se asignan aleatoriamente a una celda. Por otra parte, se asume que sólo tres empleados se seleccionan aleatoriamente y cada uno se desempeñará en todas las seis combinaciones de sistema/software. En este caso, las celdas serían homogéneas y se harían bloques no sólo en las columnas sino también en las filas. Esto es lo que se denomina un *diseño de bloques* en análisis factorial. Los autores se concentrarán en la configuración aleatoria.

El análisis factorial tiene la ventaja de ser menos costoso. Se pueden estudiar dos factores con un experimento en lugar de realizar dos pruebas independientes. Por tanto, no sólo se puede ahorrar tiempo y esfuerzo, sino que se puede lograr el mismo grado de exactitud utilizando menos unidades experimentales, ya que ambas pruebas se realizan al mismo tiempo.

O quizá de mayor valor aún es el hecho de que al utilizar el diseño factorial se puede identificar toda interacción que pueda existir entre los dos factores. Esta interacción sería imposible detectarla si los experimentos se realizaran por separado o si cada celda tuviera sólo una observación.

En análisis factorial cada factor tiene varios niveles. El sistema de computación, factor A, tiene tres niveles y los paquetes de software, factor B tienen dos niveles. Se dice que la interacción existe si un nivel del factor A funciona de manera diferente (bien sea mejor o peor) con niveles diferentes del factor B. Por ejemplo, el sistema de computación 1 puede ser más (o menos) productivo con el paquete de software 1, mientras que el sistema de computación 3 puede funcionar mejor con el paquete de software 2. Por tanto, el impacto total del factor A (sistema) sobre la productividad, depende de cuál nivel del factor B (software) se utilice. Se dice que los dos factores *interactúan* en su impacto sobre la productividad.

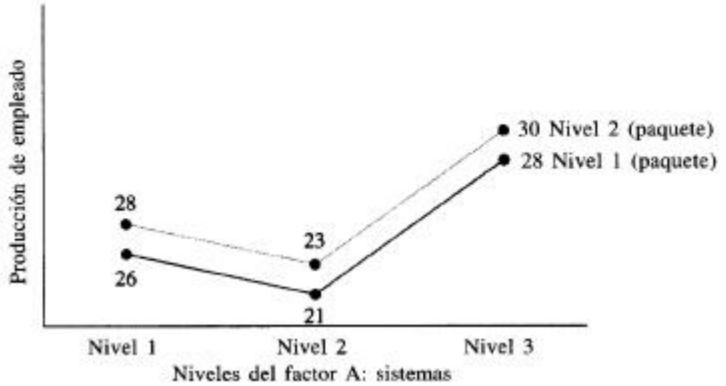
Interacción La relación entre el factor A y la variable en análisis (productividad en este caso) depende del nivel del factor B que se utilice.

La interacción puede detectarse analizando las diferencias promedio entre los niveles de un factor relativo a los niveles diferentes del otro factor. Si estas diferencias promedio son las mismas en todos los niveles de ambos factores, no existe interacción. Por otra parte, la interacción existe si las diferencias promedio entre los niveles de un factor no son las mismas para todos los niveles del otro factor.

Interacción La interacción ocurre cuando las diferencias en las respuestas promedio para los dos niveles de un factor no son las mismas en todos los niveles del segundo factor.

Esto puede mostrarse utilizando un sistema de gráficos. Se comienza con un diseño factorial que *no* presenta interacción. Se consideran nuevamente los datos de la tabla 10.5. Las medias de todas las celdas se muestran en la figura 10.2. Vale la pena destacar que las rectas no se intersectan. El nivel 2 del factor: paquete es mayor que el

Figura 10.2
Análisis factorial sin interacción



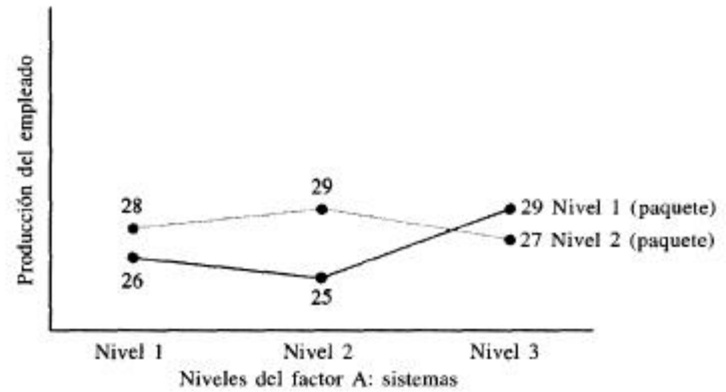
nivel 1 para todos los tres niveles del factor: sistema. De hecho, el nivel 2 del paquete agrega dos unidades al nivel 1 para todos los tres niveles del sistema de factores. Una gráfica con segmentos de recta paralela como éstas, prueba la **ausencia** de interacción. Se dice que los efectos son **aditivos**.

Se asume que las pruebas arrojaron los resultados que aparecen en la tabla 10.6. El gráfico de las medias de las celdas aparece en la figura 10.3. Vale la pena destacar que las rectas sí se cruzan. No hay una adición constante entre los niveles del factor paquete a lo largo de todos los niveles del factor sistema. Mientras que el nivel 2 del factor paquete adiciona 2 unidades al primer nivel del factor sistema (28 > 26), este patrón no continúa. De hecho, el nivel 1 del factor paquete varió de bajo a alto dependiendo de con cuál nivel del sistema se formó la pareja. Por ejemplo, se podría decir que la producción se benefició de la interacción derivada de la conformación del par del nivel 3 del factor sistema y el nivel 2 del factor paquete.

Tabla 10.6
Análisis factorial con interacción

		FACTOR A (sistema de computación) Tres niveles del factor A		
		1	2	3
FACTOR B (paquete de software)	1	28 26 24 $\mu_{11} = 26$	25 23 27 $\mu_{12} = 25$	30 29 28 $\mu_{11} = 28$
Dos niveles del factor B	2	28 27 29 $\mu_{11} = 28$	29 31 27 $\mu_{11} = 29$	28 27 26 $\mu_{11} = 27$

Figura 10.3
Análisis factorial que evidencia interacción



Como se dijo anteriormente, la ausencia de interacción se comprueba con las rectas paralelas. Sin embargo, las rectas no necesitan ser perfectamente paralelas como para concluir que la interacción no está presente. Para ilustrar esta idea, cuando se prueba la hipótesis de que $\mu = 100$, no se necesita una media muestral de exactamente 100 para no rechazar la hipótesis nula. Los valores cercanos a 100 son suficientes para el no rechazo. De la misma forma, las gráficas con las rectas que son aproximadamente paralelas, proporcionan apoyo a la hipótesis nula.

Un método más preciso para detectar la interacción reside en la prueba de hipótesis. Se pueden identificar tres hipótesis que se deben probar.

H_0: Las medias de la columna son iguales (Prueba de efectos principales del factor sistema)

H_0: Las medias de la fila son iguales (Prueba de efectos principales para el factor paquete)

H_0: No hay interacción presente

Las hipótesis alternativas para cada prueba se plantean al contrario, como es el caso usual.

De la misma manera que con el análisis de varianza a una vía y a dos vías, ahora se está preparado para desglosar de las sumas de cuadrados y construir la tabla factorial para probar la hipótesis. Los cálculos necesarios son muy tediosos cuando se realizan manualmente; entonces, para realizar esta tarea, se confiará exclusivamente en el análisis por computador.

Dados los datos de la tabla 10.5, Minitab se utiliza para generar la tabla factorial de la pantalla 10.3. La prueba para la interacción es $F = 0.00/2.83 = 0 < F_{0.05,2,12} = 3.89$. No se rechaza la hipótesis de no interacción. El impacto del sistema de computación sobre la productividad no depende de cuál programa de software se utilice. La prueba para estimar la diferencia en software es $F = 18.00/2.83 = 6.36 > F_{0.05,1,12} = 4.75$. Se rechaza la hipótesis nula de que las medias de la fila son iguales. Finalmente, la prueba para hallar una diferencia en las medias del sistema es $F = 78.00/2.83 = 27.56 > F_{0.05,2,12} = 3.89$ y se rechaza la hipótesis nula.

Pantalla 10.3

Two-way Analysis of Variance (Análisis de varianza a dos vías)

Analysis of Variance for Product (Análisis de varianza para producto)

Source (Fuente)	DF	SS	MS
Software	1	18.00	18.00
System (Sistema)	2	156.00	78.00
Interaction (Interacción)	2	0.00	0.00
Error	12	34.00	2.83
Total	17	208.00	

DF = Grados de libertad
SS = Sumas de cuadrados
MS = Cuadrado medio

MTB >

Por otra parte, se asume que la prueba de los sistemas de computación y los paquetes de software produjeron los datos que aparecen en la tabla 10.6. La impresión resultante en Minitab, pantalla 10.4, muestra que la interacción sí está presente. El valor F de $14.0/2.50 = 5.60 > F_{0.05,2,12} = 3.89$.

Pantalla 10.4

Two-way Analysis of Variance (Análisis de varianza a dos vías)

Analysis of Variance for Product (Análisis de varianza para producto)

Source	DF	SS	MS
Software	1	8.00	8.00
System	2	4.00	2.00
Interaction	2	28.00	14.00
Error	12	30.00	2.50
Total	17	70.00	

MTB >

10.6 Diseño en cuadrado latino

Otro tipo de diseño de bloques permite hacer bloques en dos variables extrañas al mismo tiempo. Vale la pena recordar el ejemplo de los sistemas de computación en el cual se deseaba bloquear el nivel de experiencia para retirar la influencia exterior. Se asume ahora que también se desea corregir la hora del día en la cual se realiza el trabajo, debido a que se conoce que la productividad de los empleados varía durante el día. Algunos empleados son más efectivos durante la mañana, otros durante la tarde, y otros son "gente nocturna". Ahora se tienen dos variables sobre las cuales se desea hacer bloques. Se quiere eliminar la influencia de dos elementos exteriores para capturar una verdadera y precisa impresión de la capacidad del sistema real. El esquema experimental apropiado es el **diseño en cuadrado latino**.

Una de las variables del bloque se asigna a las filas y la segunda a las columnas. Los tratamientos se organizan para que ocurran una vez en cada fila y en cada columna. Por consiguiente, el número de filas, el número de columnas y el número de tratamientos deben ser todos el mismo.

Diseño en cuadrado latino Se utiliza cuando se desea bloquear el efecto exterior de dos variables que pueden producir resultados equívocos.

El diseño en cuadrado latino permite al investigador obtener más información con una muestra más pequeña, ya que elimina la variación exterior haciendo el bloqueo en dos variables. A diferencia del análisis factorial, el diseño en cuadrado latino contiene sólo un elemento por tratamiento y por celda.

Como se acaba de expresar, si hay r tratamientos en consideración, se debe tener r niveles para cada una de las variables de bloqueo. Por consiguiente, se tiene $r \times r = r^2$ elementos, lo que sugiere el término *cuadrado*. El diseño se denomina *cuadrado latino* porque las letras latinas como A, B y C se utilizan para denotar los tratamientos.

Debido a que hay tres sistemas de computación que se desea probar, ahora se seleccionarán tres empleados para que trabajen en cada sistema durante cada uno de los tres períodos. Se tienen tres tratamientos (sistemas de computación) y tres niveles para cada bloque: es decir, tres niveles de experiencia y tres períodos de tiempo. Se dice que se tiene cuadrado latino 3 X 3. Los tres sistemas de computación (tratamientos) se identificarán como A, B, y C. Cada tratamiento aparecerá en todas las filas y columnas del cuadrado.

Ahora se debe dividir las sumas de los cuadrados entre las filas, las columnas y los tratamientos. Esto se logra utilizando el siguiente conjunto de fórmulas. La suma de cuadrados del bloque de filas (*SCBF*) se halla:

Suma de cuadrados del bloque de filas

$$SCBF = \frac{\Sigma(\text{ suma de fila })^2}{r} - \frac{(\Sigma X_i)^2}{r^2} \qquad [10.17]$$

en donde (suma de fila) es la suma de cada fila
r es el número de tratamientos
X_i es cada una de las observaciones

La suma de los cuadrados del bloque de columnas (*SCBC*) es:

Suma de cuadrados del bloque de columnas

$$SCBC = \frac{\Sigma(\text{ suma de columna})^2}{r} - \frac{(\Sigma X_i)^2}{r^2} \qquad [10.18]$$

en donde (suma de columna) es la suma de cada columna

La suma de cuadrados del tratamiento (*SCTR*) es

Suma de cuadrados del tratamiento

$$SCTR = \frac{\Sigma(\text{suma } TRT_i)^2}{r} - \frac{(\Sigma X_i)^2}{r^2} \qquad [10.19]$$

en donde (suma TRT_i) es la suma de cada uno de los tratamientos A, B y C.

La suma de cuadrados total es:

Suma de cuadrados total	$SCT = \Sigma(X_i)^2 - \frac{(\Sigma X_i)^2}{r^2}$	[10.20]

donde $(X_i)^2$ es el cuadrado de los nueve valores de producción. Por último, para la suma del cuadrado del error se tiene que:

Suma del cuadrado del error	$SCE = SCT - SCTR - SCBC - SCBF$	[10.21]

Se asume que el experimento produjo los datos que aparecen en la tabla 10.7. La primera celda de B/15 significa que el primer empleado que trabaja en la mañana produce 15 unidades de producción en la máquina B. La celda superior derecha de C/11 significa que el primer empleado que trabaja en la noche produjo 11 unidades de producción en la máquina C.

Tabla 10.7

Datos para un diseño latino

	Hora			
Empleado	**Mañana**	**Tarde**	**Noche**	**Total filas**
1	B/15	A/18	C/11	44
2	C/12	B/20	A/9	41
3	A/17	C/19	B/10	46
Total columnas	44	57	30	131

$\Sigma A = 44$

$\Sigma B = 45$

$\Sigma C = 42$

Aplicando las fórmulas (10.17) a (10.21) se tiene que:

$$SCBF = \frac{(44)^2 + (41)^2 + (46)^2}{3} - \frac{(131)^2}{3^2}$$
$$= 4.222$$
$$SCBC = \frac{(44)^2 + (57)^2 + (30)^2}{3} - \frac{(131)^2}{3^2}$$
$$= 121.5556$$
$$SCTR = \frac{(44)^2 + (45)^2 + (42)^2}{3} - \frac{(131)^2}{3^2}$$
$$= 1.5556$$
$$SCT = (15)^2 + (12)^2 + (17)^2 + \cdots + (10)^2 - \frac{(131)^2}{3^2}$$
$$= 138.222$$

Por tanto, $SCE = 10.89$. La tabla aparecería como la tabla 10.8.

Tabla 10.8
Diseño en cuadrado latino para la prueba de computador

Fuente de variación	Suma de cuadrados	Grados de libertad	Cuadrado medio	Valor F
Bloques de fila	4,222	2	2.111	0.3877
Bloques de columna	121,555	2	60.778	11.1621
Tratamientos	1,555	2	0.7775	0.1428
Error	10,890	2	5.445	
Total	138,222			

$H_0: \mu_1 = \mu_2 = \mu_3$ Las producciones promedio de todos los tres computadores son iguales
$F = 0.1428 < F_{0.005,2,2} = 19$; No rechazar la hipótesis nula

Fijando α al 5%, el valor F para $F_{0.05,2,2} = 19 > 0.1428$, y por tanto no se rechaza la hipótesis nula de que no hay diferencia en la producción promedio de los computadores después de ajustar la experiencia del empleado y la hora del día.

Ejercicios de la sección

15. Un productor de alambre metálico desea comparar la resistencia a la tensión utilizando tres mezclas diferentes de químicos: A, B y C. Es necesario controlar el tipo de horno utilizado para "fundir" la mezcla, y la temperatura a la cual se fundió. Utilizando los siguientes datos, ¿qué conclusión saca usted para el fabricante? Sea α 1%.

	Temperatura		
Horno	Bajo	Medio	Alto
1	A/40	B/42	C/18
2	B/70	C/19	A/45
3	C/20	A/51	B/27

16. Como director de mercadeo, usted está interesado en comparar los ingresos de tres marcas — "bueno", mejor", "óptimo"— respecto a los tenedores eléctricos que vende su empresa. Para lograrlo, usted desea corregir el área del país en la cual se encuentra ubicada la tienda, y el tipo de tienda en la cual se hizo la venta. Su asistente recolecta datos para las ventas mensuales en cientos de dólares que aparecen a continuación, pero no tiene idea sobre qué hacer con ellos. Haga la prueba usted mismo e informe al asistente sobre sus hallazgos. Sea α 1%.

	Área		
Tienda	Noreste	Sureste	Centro
Descuentos	Buena/4.2	Mejor/9.0	La mejor/12.9
Materiales	Mejor/7.3	La mejor/11.1	Buena/11.3
Centro comercial	La mejor/8.0	Buena/9.4	Mejor/10.7

17. Un investigador recolecta datos sobre los sueldos en la facultad para determinar si existe diferencia o no en los ingresos promedio de los que están en las áreas de administración, ciencias sociales y ciencias naturales. Debe eliminar los efectos extraños del rango y tamaño de la escuela. Utilizando la información que se observa aquí respecto a los salarios en miles de dólares, ¿cuáles supone usted que sean los resultados? Sea α 1% e interprete.

	Tamaño		
Rango	Pequeño	Mediano	Grande
Profesor auxiliar	Bus/65	SS/60	NS/78
Profesor asociado	SS/72	NS/81	Bus/79
Profesor titular	NS/82	Bus/73	SS/79

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:

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

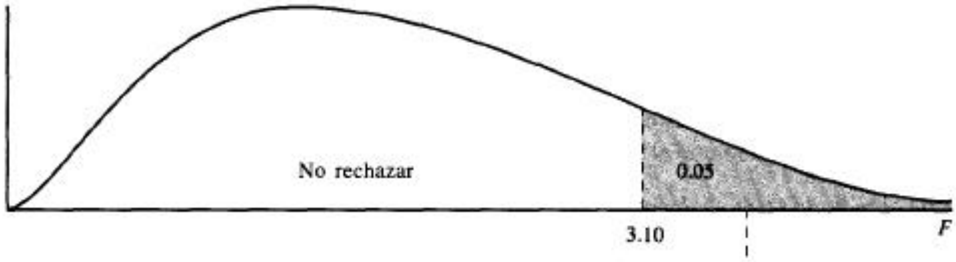
SCE es:

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

Entonces,

$$CMTR = \frac{CSTR}{c-1}$$
$$= \frac{48,023}{4-1}$$
$$= 16,008$$
$$CME = \frac{SCE}{n-c}$$
$$= \frac{35,492}{24-4}$$
$$= 1,775$$

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.

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

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 $\bar{X}_1$ 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**	$\bar{X}_i$
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
$\bar{X}_j$	51.3	38.3	47.1	40.2	
$\bar{\bar{X}} = 44.23$					

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:

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

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

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.

4. **Conducción en Circle** Circle Trucking Company desea determinar si el tiempo promedio de conducción es el mismo para cada una de las tres rutas. El director de tránsito para Circle considera que es necesario corregir las condiciones climáticas así como la pericia de los conductores. Se identifican tres niveles de condiciones climáticas: deficiente, regular y buena. Se seleccionan tres conductores con capacidades cambiantes y cada uno cubre tres rutas bajo cada una de las tres condiciones climáticas. Los resultados se reportan en el diseño en cuadrado latino que se muestra en el texto. Vale la pena recordar que las letras latinas indican las variables en consideración, en este caso, las rutas. Se registran los tiempos en minutos.

	Clima			
Conductor	**Deficiente**	**Regular**	**Bueno**	**Total filas**
1	A/20	C/18	B/17	55
2	C/22	B/10	A/10	42
3	B/18	A/9	C/8	35
Total columnas	60	37	35	132
	ΣA = 39	ΣB = 45	ΣC = 48	

Solución

Utilizando el diseño en cuadrado latino hay dos variables extrañas para corregir (destrezas y clima), los cálculos aparecen como:

$$SCF = \frac{(55)^2 + (42)^2 + (35)^2}{3} - \frac{(132)^2}{9}$$
$$= 68.667$$
$$SCC = \frac{(60)^2 + (37)^2 + (35)^2}{3} - \frac{(132)^2}{9}$$
$$= 128.667$$
$$SCTR = \frac{(39)^2 + (45)^2 + (48)^2}{3} - \frac{(132)^2}{9}$$
$$= 14$$
$$SCT = (20)^2 + (22)^2 + \cdots + (8)^2 - \frac{(132)^2}{9}$$
$$= 230$$
$$SCE = 230 - 14 - 128.667 - 68.667$$
$$= 18.663$$

Fuente de variación	Suma de cuadrados	Grados de libertad	Cuadrado medio	Valor F
Bloques de filas (conductor)	686,667	2	34.33	3.679
Bloques de columna (clima)	128,667	2	64.33	6.895
Tratamiento (ruta)	14,000	2	7.00	0.750
Error	18,663	2	9.33	
Total	230,000			

Si la hipótesis va a probarse al nivel del 5%, $F_{0.05,2,2} = 19 > 0.750$. Por tanto, Circle no debería rechazar la hipótesis nula.

Lista de fórmulas

[10.1] $$\overline{\overline{X}} = \frac{\Sigma X_{ij}}{n}$$ Gran media de todas las observaciones del experimento

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

[10.3] $$SCT = \sum_{i=1}^{r}\sum_{j=1}^{c}(X_{ij} - \overline{\overline{X}})^2$$ La suma de cuadrados total es la variación de los valores alrededor de $\overline{\overline{X}}$

[10.4] $$SCTR = \Sigma r_j(\overline{X}_j - \overline{\overline{X}})^2$$ La suma de los cuadrados del tratamiento mide la variación en las medias de la columna alrededor de la gran media

[10.5] $$SCE = \Sigma\Sigma(X_{ij} - \overline{X}_j)^2$$ 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
[10.17]	$SCF = \frac{\Sigma(\text{suma fila})^2}{r} - \frac{(\Sigma X_i)^2}{r^2}$	Suma de cuadrados del bloque de filas para el diseño en cuadrado latino
[10.18]	$SCC = \frac{\Sigma(\text{suma columna})^2}{r} = \frac{(\Sigma X_i)^2}{r^2}$	Suma de cuadrados del bloque de columnas para el diseño en cuadrado latino
[10.19]	$SCTR = \frac{\Sigma(\text{suma } TRT)^2}{r} - \frac{(\Sigma X_i)^2}{r^2}$	Suma de cuadrados del tratamiento para el diseño en cuadrado latino
[10.20]	$SCT = \Sigma(X_i^2) - \frac{(\Sigma X_i)^2}{r^2}$	Suma total de cuadrados para el diseño en cuadrado latino
[10.21]	$SCE = SCT - SCTR - SCBC - SCBF$	Suma de cuadrados del error para el diseño en cuadrado latino

Ejercicios del capítulo

18. Como analista de mercadeo usted desea determinar si hay una diferencia en las ventas promedio en tres mercados: Pittsburgh, Toledo y Columbus. ¿Por qué no eliminar la necesidad del análisis de varianza tan sólo probando para hallar la igualdad entre cada par de mercados? ¿Si no se rechaza cada hipótesis, usted puede concluir entonces que todos los tres mercados tienen la misma media?

19. ¿Por qué un valor *F* alto resulta en un rechazo de la hipótesis nula de medias iguales?

20. ¿Cuál es la diferencia entre un modelo de efectos fijos y un modelo de efectos aleatorios?

21. ¿Cuál es la diferencia entre un análisis de varianza a una vía y un análisis de varianza a dos vías?

22. ¿Bajo qué condiciones utilizaría usted el análisis factorial? ¿Un diseño en cuadrado latino?

23. ¿Cuál es la diferencia entre un diseño balanceado y un diseño no balanceado?

24. De acuerdo con un artículo en la revista *Fortune*, las empresas manufactureras más pequeñas tienen cada vez más dificultad para recibir los pedidos de sus proveedores dentro de un tiempo razonable. A medida que la economía se calienta y la capacidad de producción se restringe, los pedidos tienden a retrasarse. Como supervisor de producción de Novelties, Inc., usted desea probar el tiempo de entrega promedio en días, para los pedidos que usted hace con tres proveedores diferentes del componente del cojín de lujo whoopie que fabrica su empresa. Aquí se presentan los tiempos de entrega. A un nivel del 5%, ¿existe una diferencia en los tiempos promedio?

Proveedor 1: 5, 6, 6, 5, 6, 6, 7
Proveedor 2: 5, 4, 5, 5, 6, 5, 4
Proveedor 3: 4, 5, 2, 6, 5, 2, 4

25. Dados los resultados del problema anterior, ¿cuál(es) proveedor(es) recomendaría? ¿Cuál(es) recomendaría evitar? Sea $\alpha = 0.01$. Resuma los resultados con el subrayado común.

26. Una discusión en *American Agriculture* llamó la atención sobre el efecto de diferentes suplementos alimenticios en grano en las tasas de crecimiento de los pollos criados comercialmente. En Charlie's Chicken Ranch, se practicó una prueba en la cual 18 pollos se dividieron de forma homogénea en tres grupos, y cada grupo era alimentado con un suplemento en particular. Los incrementos en crecimiento durante un período de seis semanas, medidos en libras, aparecen en la siguiente tabla. A un nivel del 10%, ¿parece haber evidencia que indique cuál suplemento debería utilizar Charlie en el futuro?

	Suplemento		
Pollo	**Grow-big**	**Cluckers Choice**	**Cock of the Walk**
1	2.2	3.7	3.8
2	2.4	2.1	4.1
3	2.7	3.2	3.9
4	3.8	2.9	2.7
5	3.2	3.9	4.1
6	3.9	3.8	3.2

27. Hay muchos aspectos por considerar en el desarrollo de una estrategia de mercadeo. La ubicación de la tienda es una preocupación importante. PDQ, una cadena de materiales en todo el sudeste, informó los resultados en una publicación interna sobre una encuesta de ingresos semanales provenientes de tiendas con ubicaciones urbanas, suburbanas y rurales. Los datos, los cuales se han simplificado para nuestros propósitos, se muestran a continuación. ¿Puede llegarse a alguna conclusión respecto a las ubicaciones principales para las tiendas? Determine cuál(es) ubicación(es), si la(s) hay, es (son) mejor(es). Sea $\alpha = 0.05$. Muestre los resultados con subrayado común.

	Ubicación		
Tienda	**Urbana**	**Suburbana**	**Rural**
1	789	612	718
2	762	655	655
3	722	725	725
4	745	609	645
5	802	632	622

28. Una emisión reciente de *Bicycling* discutía el uso de programas de computación en el desarrollo de un régimen de entrenamiento. Uno de esos programas de computación probaba a varios ciclistas que estaban en condición física superior y concluyeron que para que los entrenamientos extendidos fueran más benéficos deberían hacerse al 60 o 70% de la proporción máxima de latidos del corazón del individuo (aproximadamente 220 latidos por minuto menos la edad de la persona). Entrenamientos más intensos de una duración más corta deberían llegar a 80 o 90% de ese máximo.

Los tres programas de entrenamiento fueron diseñados para determinar las técnicas óptimas de entrenamiento. Se asume que los individuos aplicaron cada programa, y después de seis semanas, se monitoreó la tasa de latidos finales. Los datos se encuentran registrados en la siguiente tabla y representan los porcentajes de las tasas máximas recomendadas. A un nivel del 5%, parece haber diferencia en las tasas máximas promedio de los latidos del corazón?

	Programa de entrenamiento		
Ciclista	**1**	**2**	**3**
1	0.62	0.68	0.72
2	0.73	0.52	0.69
3	0.59	0.59	0.73
4	0.82	0.63	0.74
5	0.79	0.61	0.68

29. Complete los cálculos del problema anterior, fijando $\alpha = 0.10$. Haga gráficas para las distribuciones F en cada caso, mostrando una comparación de las pruebas para cada valor α.

30. *Business Week* manifestó que John F. Akers, director ejecutivo de la IBM, dijo que consideraba improbable que en el futuro cercano, el crecimiento anual en ventas de la IBM del 6.1% pudiera seguir el ritmo del crecimiento general del sector, cifrado en el 9.2% de toda la industria. Este rezago en los recibos se debía en parte a la confianza de la IBM en los macrocomputadores, cuyo mercado en las ventas mundiales había bajado al tercer lugar, tras los PCs y los minicomputadores.

Los datos trimestrales sobre el porcentaje del incremento en las ventas para cinco períodos se han recolectado para cada uno de estos mercados de hardware. Los resultados son

Macrocomputadores: 3.2, 4.8, 4.1, 4.2, 3.9

PCs: 8.7, 9.2, 9.3, 8.3, 8.9

Minicomputadores: 9.1, 9.4, 8.7, 9.5, 9.9

¿Estos datos muestran alguna diferencia significativa en los incrementos promedio en ventas a un nivel del 1%?

31. *USA Today* imprimió una historia sobre el uso de detectives privados para efectos de descubrir algunos hechos que pudieran hacer que la empresa fuera menos deseable desde el punto de vista de una fusión por adquisición. "M & A work" dice J.B. Kroll, presidente de Kroll y Associados, una empresa de detectives con sede en Nueva York "da cuenta de por lo menos el 20% de los US$50 millones que Kroll debería percibir este año": Petroquímicos, banca, computadores y electrónica son industrias especialmente fértiles para M & A .

Se asume que seis empresas de cada industria se investigan aleatoriamente mediante encuestas para determinar los montos involucrados en las propuestas de adquisición, y los resultados aparecen a continuación. ¿Podría concluirse, al nivel del 5% que existe alguna diferencia en las propuestas promedio entre estas industrias? Los valores están en millones de dólares.

Licitación	Petroquímicos	Banca	Computadores	Electrónica
1	919.3	842.7	647.3	743.7
2	874.2	1,144.7	873.2	747.3
3	832.7	942.3	714.4	812.5
4	732.9	747.1	652.8	643.7
5	893.2	812.7	855.6	682.1
6	1,321.4	855.6	642.1	632.1

32. El gran lobo malo desea engordar a los tres cerditos para la cena de la festividad. Sus datos muestran que la cantidad de alimento en libras que cada cerdito ha comido, para una muestra de cinco días, es la siguiente:

Cerdito 1	Cerdito 2	Cerdito 3
12	14	19
15	16	18
14	15	17
16	18	16
13	12	18

a. ¿Al nivel del 5%, existe alguna diferencia en la cantidad promedio ingerida por los cerditos?

b. Utilice el criterio de Tukey para determinar cuál cerdito o cerditos deben engordarse primero. Resuma con el subrayado común.

c. Realice las mismas pruebas por pares utilizando el método *DMS*. Resuma con el subrayado común.

33. Cuando el presidente Clinton retiró el embargo comercial a Vietnam en el año de 1994, muchas compañías de Estados Unidos iniciaron relaciones comerciales. Los datos que aparecen en el cuadro son los montos comerciales (en miles de millones de dólares) para cuatro industrias. Sea α 1%. ¿Cuál es su conclusión respecto a la diferencia en los niveles comerciales promedio de estas cuatro industrias? Utilice el método de Tukey para sacar conclusiones sobre las comparaciones por pares. Resuma con el subrayado común.

Petróleo	Bebidas suaves	Maquinaria de construcción	Computadores
2.1	5.6	1.5	4.5
2.5	6.2	1.0	4.2
2.6	7.8	1.8	4.1
2.1	6.8	1.9	4.6
3.5	5.4	1.7	4.2

34. Un economista del Departamento de Estado de Estados Unidos, especialista en demografía, obtuvo los siguientes datos que demuestran una inmigración anual en millones de personas de cuatro países diferentes en una muestra de cinco años. Si α es 1%, ¿qué conclusión puede sacar respecto a la comparación de las medias?

Inglaterra	México	Canadá	Francia
1.5	2.5	0.9	0.8
1.6	2.6	1.1	0.6
1.8	2.7	0.8	0.8
1.7	2.9	1.1	0.7
1.6	2.5	0.9	0.9

35. Utilice tanto el método de Tukey como el *DMS* para hacer comparaciones por pares sobre los resultados del problema anterior. Complete con un subrayado común.

36. Una empresa de consultoría prueba tres métodos diferentes para la toma de decisiones, comparando la cantidad de tiempo que requieren los equipos administrativos que utilizan cada uno de estos métodos para terminar los proyectos financieros. Con base en los datos que aparecen a continuación, ¿parece que el método de toma de decisiones afecta el tiempo necesario para terminar el proyecto? Sea α 10%. Los tiempos se miden en días transcurridos hasta la terminación del proyecto.

	Método		
Equipo administrativo	**1**	**2**	**3**
1	40	37	43
2	31	31	32
3	27	49	43
4	52	28	44
5	63	37	32
6	57	27	37

37. Como director de publicidad de su firma, usted desea comparar la efectividad de varios formatos publicitarios. Se muestran tres anuncios a varios compradores, quienes los califican en una escala del 10 al 50. Los resultados aparecen aquí. ¿Cuáles anuncios seleccionaría, si los hay, sobre los otros, para distribuirlos en masa? Sea $\alpha = 0.10\%$. Resuma los resultados con el subrayado común.

	Anuncio publicitarios		
Comprador	**1**	**2**	**3**
1	45	40	30
2	40	30	35
3	35	30	30
4	35	35	30
5	40	40	35
6	35	25	30
7	30	25	30

38. Se practicó una encuesta informal sobre los hábitos de citas de los estudiantes en una universidad del estado. Aquí se muestran los resultados que registran el número de citas por mes. A un nivel del 5%, ¿parece haber alguna diferencia por clase en la frecuencia de las citas? Si es así, utilice tanto el método de Tukey como el *DMS* para determinar cuáles son diferentes. Resuma los resultados con subrayado común.

Estudiante	Clase Fr	So	Ju	Sr
1	2	2	3	4
2	2	0	5	2
3	1	2	6	5
4	2	6	4	3
5	0	4	3	3
6	3	4	6	4

39. Cuando la ola de ofertas de adquisiciones hostiles alcanzó un frenesí exagerado a finales de los años 80, muchas empresas reportaron el uso de "píldoras venenosas" para hacerse menos atractivos ante otras empresas que estuvieran planeando adquisiciones. Las "píldoras" eran acciones que se realizaban para desilusionarse de una adquisición, e incluía planes de deudas para pensionados, políticas de opción de compra de acciones y paracaídas dorados para ejecutivos pensionados, lo cual era desfavorable para la empresa que adquiría. Un estudio diseñado para medir los efectos comparables de estas tres acciones, registraron cambios en los precios de las acciones de varias empresas que lo utilizaron. Los datos se muestran a continuación. A un nivel del 5%, ¿parece que algunas píldoras son más efectivas para reducir el precio de las acciones de las empresas?

Empresa	Planes de préstamos para pensionados	Opción de compra de acciones	Paracaídas dorados
1	-1.55	-2.10	0.20
2	-2.54	-3.20	-1.10
3	-3.55	-1.47	1.55
4	-2.10	1.01	-1.25
5	1.50	-3.55	2.10
6	-2.17	-2.99	1.20

40. Utilizando los datos del problema anterior, ¿cuáles píldoras recomendaría usted a su junta directiva si desearan reducir los precios de las acciones para hacer su empresa menos atractiva? Sea $\alpha = 0.05$. Muestre el subrayado común apropiado.

41. Una empresa nacional que comercializa una loción bronceadora selecciona aleatoriamente a cinco personas para que prueben las tres fórmulas. Cada loción se aplica a diferentes partes del cuerpo en todos los sujetos de prueba. Después del tiempo designado en el sol, el factor de bronceado se mide utilizando una escala desarrollada por un dermatólogo reconocido.

Sujetos de prueba	Fórmulas de bronceado *Tan Your Hide*	*Burn Not*	*Tanfastic*
1	3	4	5
2	5	4	4
3	4	3	4
4	4	5	3
5	3	2	4

Sea $\alpha = 0.01$ y determine si alguna fórmula promociona un bronceado más que las otras. Si es así, identifique cuáles son. Dadas las diferencias en los sujetos de prueba en la capacidad bronceadora natural, haga la prueba para determinar si se necesita el bloqueo.

42. Una empresa de taxis intenta crear un sistema de rutas que minimizará el tiempo que se pasa manejando a ciertas localidades. Hay cuatro rutas en consideración. Usted es contratado como consultor estadístico para que ayude. Cinco taxistas utilizan cada una de las rutas y registran sus tiempos en minutos. A un nivel del 5%, ¿puede identificar cuál(es) ruta(s) es(son) más rápida(s)? ¿Debería bloquear sobre el conductor? ¿Cuál(es) ruta(s) aconsejaría a la compañía que utilizara?

	Ruta			
Taxista	**1**	**2**	**3**	**4**
1	12	15	17	13
2	18	18	18	17
3	10	11	15	9
4	13	12	12	15
5	18	14	12	15

43. Cuatro inspectores de control de calidad acaban de terminar un curso de entrenamiento. Cada uno inspecciona las mismas cinco unidades de producto terminado y se le pide calificarlas. A un nivel del 5%, parece que los inspectores asignan las mismas calificaciones promedio?

	Inspector			
Producto	**Fred**	**Sam**	**Terri**	**Minerva**
1	55	45	52	51
2	87	86	78	82
3	69	68	67	62
4	75	71	72	68
5	69	71	75	68

44. Un grupo de consumidores locales está interesado en determinar si las tiendas en diferentes localidades de toda la ciudad ofrecen los mismos precios para los alimentos. A una canasta de muestra con cinco artículos se le fijó precio en los tres sitios de la ciudad. A un nivel del 1%, ¿parece que los precios en promedio son los mismos para los tres lugares?

	Lugar		
Producto	**Centro**	**Suburbio**	**Afueras**
1	2.92	3.15	4.35
2	1.21	1.32	1.10
3	5.12	4.79	5.20
4	0.98	1.00	0.98
5	1.50	1.45	1.50

45. El propietario de Stop & Shop ha probado cuatro políticas diferentes de cambios de cheques para reducir el gran número de cheques sin fondos que recibe su tienda. Él desea saber cuál política, si la hay, minimiza el problema. Debido a que el área del centro en la cual se encuentra ubicada la tienda puede afectar el número de cheques sin fondos, se mantienen los registros diarios sobre el número de cheques devueltos por fondos insuficientes bajo cada política en todas las cinco tiendas. Los resultados se muestran a continuación. Sea α 1%. ¿Es una política mejor que otra?

	Política			
Lugar	**A**	**B**	**C**	**D**
1	22	35	47	43
2	27	28	31	47
3	20	17	15	12
4	18	20	23	17
5	15	18	18	19

46. D. Bumstead trabaja como agente de publicidad para Dithers & Company. El Sr. Bumstead debe determinar si tres anuncios publicitarios diferentes evocan la misma respuesta de los televidentes. El señor Bumstead, sabiendo que la hora del día puede influir en las actitudes, decide sabiamente probar cada anuncio en los mismos tres períodos durante períodos de 24 horas. ¿Los resultados que aparecen a continuación dicen al señor Bumstead si existe una diferencia en las calificaciones promedio de los televidentes? Se asume un valor α del 1%.

	Anuncio		
Período	**A**	**B**	**C**
Mañana	10	15	12
Tarde	2	2	3
Noche	2	1	2

47. Bantam Books utiliza tres técnicas de impresión diferentes. Un estudio en control de calidad encontró excesivos errores de impresión como tachones, sobreimpresión, caracteres borrosos y algunas páginas incluso quedaron en blanco. Para determinar si hay alguna diferencia en el número promedio de errores con base en el método de impresión, se imprimió un cierto pasaje utilizando cada método en los cuatro tipos de papel que utiliza Bantam. ¿Los resultados indican que uno o más métodos son mejores? Sea α 5%.

	Método de impresión		
Tipo de papel	**A**	**B**	**C**
W	2	1	1
X	3	3	2
Y	5	6	3
Z	4	4	4

Ejercicio por computador

Dada su considerable experiencia en la industria de inversión, usted decide establecer su empresa de corretaje. Su primera tarea es comparar los rendimientos promedio de tres industrias diferentes. Ingrese al archivo ANOVA de su disco de datos. La primera columna de datos ("Rate") es la tasa de crecimiento para varias compañías en las tres industrias que usted desea examinar. La segunda columna ("Ind") identifica cada industria y está codificada como 1 para industria electrónica, 2 para industria de seguros, y 3 para industria bancaria.

Usted también desea comparar los rendimientos promedio de diferentes inversionistas con base en el tamaño de su inversión. La tercera columna del conjunto de datos es la tasa de rendimiento TR o rate of return "ROR" que los inversionistas ganaron durante el año anterior. La cuarta columna identifica el tamaño de la inversión como pequeña, mediana y grande codificado como 1, 2, y 3 respectivamente. Usted también considera necesario corregir el nivel de actividad en la cuenta, en la comparación de los rendimientos promedio de inversiones de diferente tamaño. Por tanto, la quinta columna está codificada de 1 a 5 para diferenciar entre las cuentas comerciales con menos actividad y las que presentan un nivel mayor de actividad. Entre más alto sea el código de valor, la cuenta se comercializa más activamente.

Prepare todo el análisis necesario y presente sus descubrimientos en forma de reporte estadístico, como se presenta en el apéndice I.

PUESTA EN ESCENA

John S. Mill recibió recientemente un título de maestría en economía orientado hacia la investigación y durante los últimos 18 meses ha estado trabajando con la CIA (Agencia Central de Inteligencia) como analista económico. Acompañó a la delegación de Estados Unidos a Denver para las charlas de los G-7 que se mencionaron en la sección "Escenario" a comienzos de este capítulo. Su proyecto actual es identificar características que puedan servir como predictores de disturbios dentro de un país extranjero. Una vez que se hayan aislado tales factores, los países que presentan esas tendencias serán objeto de estudio y análisis adicional.

El antecedente económico de John le ha enseñado que las naciones con alto índice de desempleo y otras condiciones económicas desfavorables, con frecuencia tienen más probabilidad de sufrir desórdenes. La división de análisis de datos de la CIA proporciona a John cifras sobre desempleo y tasas de inflación para varios países en cuatro áreas de todo el mundo. Estos valores aparecen aquí primero con la tasa de desempleo y después con la tasa de inflación. Así, una entrada como "17.5% & 20.2%" significa que el país tiene una tasa de desempleo de 17.5 % y una tasa de inflación de 20.2%.

La agencia ha presupuestado US$4.700.000 para la adquisición de equipo adicional de recolección de información de inteligencia durante los próximos años. John debe preparar un informe indicando cómo deberían asignarse estos fondos. ¿Cómo puede utilizar John estos datos en el proceso de toma de decisiones respecto a la asignación apropiada del dinero? Los superiores de John lo han prevenido de que a toda costa, debe evitar rechazar la hipótesis de diferencias no significantes en estos países, incluso si no existe ninguna.

País	Medio Oriente	Este de Europa	Lejano Oriente	Centroamérica
1	17.5% & 20.2%	10.2% & 8.2%	18.1% & 9.7%	27.8% & 50.2%
2	16.1% & 10.2%	8.4% & 7.3%	30.2% & 17.5%	39.7% & 40.3%
3	12.5% & 8.7%	7.2% & 6.3%	25.3% & 21.2%	37.8% & 47.3%
4	15.2% & 17.1%	7.5% & 5.3%	19.7% & 10.2%	42.1% & 80.7%
5	22.3% & 18.7%	9.2% & 6.1%	21.3% & –	37.0% & 38.9%
6	18.3% & 23.7%	10.7% & –	22.0% & –	48.0% & 73.1%
7	19.2% & –	17.1% & –	24.1% & –	38.7% & 63.2%

Del escenario a la vida real

Sí, la CIA tiene un sitio público en Internet (*www.odci.gov/cia*). Cuando usted visite este sitio, necesitará hacer clic en un área de "continúe" para llegar a la página de características. Allí, seleccione "Publications", y luego seleccione "World Fact Book". En esta área, usted encontrará información similar a la utilizada en el análisis de la sección "Puesta en escena". El sitio proporciona datos sobre desempleo, inflación, y PIB por país para la mayoría de los países del globo. Al hacer clic en "all countries" o en regiones en particular, busque los datos más actuales sobre estos tres indicadores económicos para cada uno de los 10 países de la tabla que aparecen en la sección inicial del capítulo "Escenario". Se proporcionan los datos disponibles en este sitio dos años atrás al tiempo actual y sólo se suministran los datos de un año.

De manera alterna, se encuentran disponibles los datos internacionales sobre los indicadores económicos claves que están actualizados y proporcionan los datos de los tres últimos años sobre el Departamento de Estado de Estados Unidos (*www.state.gov*). Esta información económica resumida, forma parte de los informes completos de investigación comercial sobre los países individualmente. Para su escogencia de tres de los 10 países enumerados en el inicio del capítulo "Escenario" busque los datos económicos resumidos proporcionados en este sitio. En la página de presentación (Home Page) de U.S. State Department (Departamento de Estado de Estados Unidos) seleccione "Services". Baje hasta el encabezado Business, y después haga clic en "Country Commercial Guides" (Guías comerciales del país). Luego seleccione una nueva región geográfica y después seleccione un país de interés. En el perfil comercial del país, baje a sección X: Apéndices para buscar información resumida sobre la economía nacional. A medida que usted se mueve por este informe, vale la pena destacar otras secciones importantes de estos perfiles.

Si usted estuviera haciendo investigación internacional comparativa, ¿cuáles sitios, la CIA o el Departamento de Estado de Estados Unidos, sería su primera selección y por qué?