

Unidad 2

- Inferencia con Muestras Pequeñas

9.1 Introducción

9.2 La distribución “t de Student”

9.3 Inferencias con muestras pequeñas para la media de una población

9.4 Inferencias con muestras pequeñas para la diferencia entre dos medias

9.5 Una prueba para diferencias apareadas

9.6 Inferencias acerca de la varianza de una población

9.7 Comparación de las varianzas de dos poblaciones

9.8 Resumen

9 Inferencia con muestras pequeñas

9.1 Introducción

En el capítulo anterior se discutieron métodos de inferencia, con base en muestras grandes, acerca de la media de una población y de la diferencia de las medias de dos poblaciones; sin embargo es frecuente que el tamaño de muestra se vea limitado por el costo y por el tiempo. Cuando es éste el caso, los procedimientos con muestras grandes del capítulo 8 resultan inadecuados y se requieren otros procedimientos de estimación y de prueba de hipótesis. En este capítulo se estudian varios procedimientos basados en muestras pequeñas que están muy relacionados con los procedimientos de muestras grandes del capítulo 8. En particular se consideran en este capítulo procedimientos de estimación y pruebas de hipótesis acerca de la media de una población, la diferencia entre la media de dos poblaciones, la varianza de una población y la comparación de las varianzas de dos poblaciones. (Las pruebas acerca de proporciones con muestras pequeñas fueron discutidas ya en el capítulo 6.)

9.2 La distribución t de Student

Para entrar en materia considere el siguiente problema. Se ha realizado un experimento muy costoso para evaluar un nuevo proceso para la producción de diamantes sintéticos. Se generan seis diamantes mediante el nuevo proceso y resultan estos de 0.46, 0.61, 0.52, 0.48, 0.57 y 0.54 kilates.

Un estudio de los costos del proceso indican que el peso promedio de los diamantes debe ser mayor de 0.5 kilates para que el proceso pueda ser operado con utilidad. Debe, entonces, decidirse si los pesos observados en estos seis diamantes presentan evidencia suficiente para indicar que el peso promedio de los diamantes, producidos mediante el proceso, es mayor que 0.5 kilates. Esto es, se desea probar la hipótesis nula que $\mu = 0.5$ contra la hipótesis alternativa $\mu > 0.5$.

Como se recordará, por virtud del teorema central del límite, la estadística de prueba

$$z = \frac{\bar{y} - \mu}{\sigma / \sqrt{n}}$$

posee una distribución muestral que es aproximadamente normal cuando n es grande. Para $\alpha = .05$, puede emplearse una estadística de una sola cola y rechazar la hipótesis nula si $z > 1.645$. Todo esto suponiendo que se conoce σ o que se dispone de un buen estimador s , basado en una muestra razonablemente grande. (Se ha sugerido $n \geq 30$.) Desafortunadamente estas suposiciones no son necesariamente satisfechas si $n = 6$, como es éste el caso. ¿Cómo, entonces, se puede probar la hipótesis nula $\mu = 0.5$ contra la alternativa $\mu > 0.5$ cuando se tiene una muestra pequeña?

Problemas como el arriba expuesto no son nuevos. Han recibido la atención de estadísticos y experimentadores desde comienzos de este siglo. Si la desviación estándar muestral s substituye a σ en z , cuando el tamaño de muestra es pequeño, ¿seguirá teniendo la estadística de prueba resultante, una distribución muestral aproximadamente normal? Más específicamente, ¿puede usarse como región crítica $z > 1.645$? En este caso ¿excederán 1.645 sólo el 5% de los valores de z si el experimento se repitiese un buen número de veces y H_0 fuese cierta?

La respuesta a estas preguntas, como muchos de los problemas encontrados en las ciencias, pueden ser resueltos vía experimentación; en este caso se tomarían muestras pequeñas ($n = 6$) repetidas veces y se calcularía el valor de la estadística de prueba, para construir un histograma con los valores calculados de la estadística de prueba, para visualizar así la forma de la distribución y la localización de la región de rechazo.

La distribución de la estadística de prueba

$$t = \frac{\bar{y} - \mu}{s / \sqrt{n}}$$

para *muestras pequeñas tomadas de una población normal* fue descubierta por W.S. Gosset y publicada (en 1908) bajo el pseudónimo de "Student." El se refirió a la cantidad estudiada como t y la cantidad ha sido conocida desde entonces como **t de Student**. Se omite la complicada forma de la expresión matemática para la distribución muestral de t pero se describen algunas de sus características.

En muestreo repetido, la distribución de la estadística de prueba

$$t = \frac{\bar{y} - \mu}{s / \sqrt{n}}$$

es como en el caso de la z de forma acampanada y perfectamente simétrica con respecto a $t = 0$. A diferencia de z , es mucho más variable y las colas de la distribución de t son más “densas.” Este fenómeno es fácilmente explicable. La variabilidad de z en repetidas muestras es debida sólo a la variabilidad de $\bar{y}$; las otras cantidades que aparecen en z , que son n y σ , no son aleatorias. Por otro lado, la variabilidad de t es el resultado de dos cantidades aleatorias $\bar{y}$ y s que, puede ser demostrado, son independientes una de la otra. Cuando $\bar{y}$ es muy grande, s puede ser muy pequeña, y viceversa, resultando de esto que t sea más variable que z en muestreo repetido (ver figura 9.1). Finalmente, como se podría suponer la variabilidad de t disminuye cuando n aumenta, porque la estimación s de σ está basada en más y más información. Cuando n es infinitamente grande, ambas distribuciones, la de t y la de z , son idénticas. Es decir, como fue descubierto por Gosset, la distribución de t depende del tamaño de la muestra n .

El divisor $(n - 1)$ de la suma de cuadrados que aparece en la fórmula para s^2 es denominado el número de **grados de libertad** asociados a s^2 . El origen del término “grados de libertad” está ligado a la teoría estadística involucrada en la distribución de probabilidad de s^2 . No se hará referencia posterior a este punto. Solamente se hará notar que cuando la estadística t está basada en n observaciones se dice que la distribución posee $(n - 1)$ grados de libertad (g.l.)

Los valores críticos de t que separan la región de rechazo y la región de aceptación para las pruebas estadísticas se presentan en la tabla 4 del apéndice. La tabla 9.1 ilustra el formato y muestra algunas entradas de la tabla 4. El valor tabulado t_α es el valor de t tal que a la derecha de éste queda una área α , como se ilustra en la figura 9.2. Los grados de libertad (g.l.) asociados a s^2 aparecen en la primera y última columnas de la tabla y las t_α 's para indicar los distintos valores de α aparecen en la fila superior. Entonces si se desea, por ejemplo, encontrar el valor de t tal que el 5% del área quede a su derecha se usará la columna marcada $t_{.050}$. El valor crítico de t en el ejemplo de los diamantes que se encuentra en la columna $t_{.050}$ y en la fila correspondiente a g.l. = $n - 1 = 6 - 1 = 5$ es $t_{.050} = 2.015$. Este valor se señala en la tabla 9.1. Entonces se rechazará H_0 cuando $t > 2.015$.

Debe notarse que el valor crítico de t es siempre mayor que el correspondiente al valor crítico para z para un α dado. Por ejemplo, cuando $\alpha = .05$, el valor crítico correspondiente de $z = 1.645$. Al descender en la columna $t_{.05}$ se

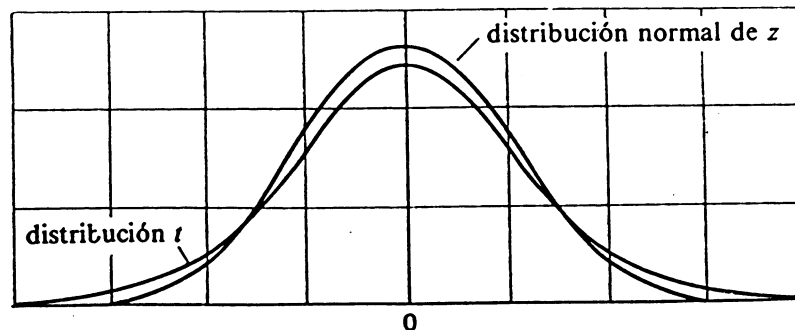
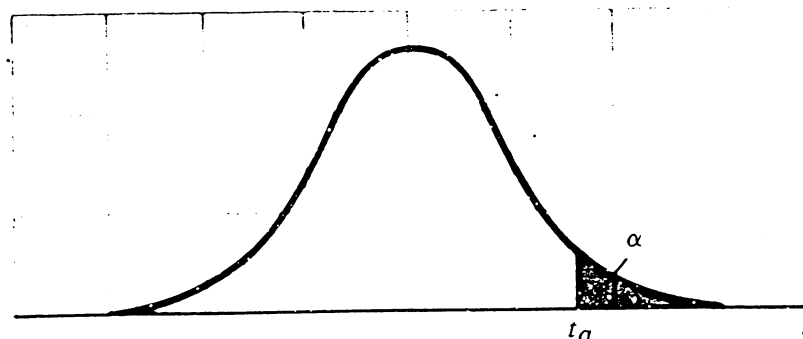


Figura 9.1 La distribución normal de z y la distribución t basada en $n = 6$ observaciones (5 g.l.)

Figura 9.2 Valores tabulados de la t de Student

observa que los valores críticos para t decrecen, reflejando el efecto de un mayor tamaño de muestra en la estimación de σ . Finalmente cuando n es infinitamente grande el valor crítico de t es igual a 1.645.

Tabla 9.1 Formato de la tabla de la t de Student, tabla 4 del apéndice

g.l.	$t_{.100}$	$t_{.050}$	$t_{.025}$	$t_{.010}$	$t_{.005}$	g.l.
1	3.078	6.314	12.706	31.821	63.657	1
2	1.886	2.920	4.303	6.965	9.925	2
3	1.638	2.353	3.182	4.541	5.841	3
4	1.533	2.132	2.776	3.747	4.604	4
5	1.476	2.015	2.571	3.365	4.032	5
6	1.440	1.943	2.447	3.143	3.707	6
7	1.415	1.895	2.365	2.998	3.499	7
8	1.397	1.860	2.306	2.896	3.355	8
⋮	⋮	⋮	⋮	⋮	⋮	⋮
27	1.314	1.703	2.052	2.473	2.771	27
28	1.313	1.701	2.048	2.467	2.763	28
29	1.311	1.699	2.045	2.462	2.756	29
inf.	1.282	1.645	1.960	2.326	2.576	inf.

La razón de haber elegido $n = 30$ (elección que es arbitraria) como la línea divisoria entre muestras grandes y pequeñas se justifica al examinar la tabla 9.1. Para $n = 30$ el valor crítico de $t_{.05} = 1.699$ es bastante cercano numéricamente a $z_{.05} = 1.645$. Para una prueba de dos colas, basada en $n = 30$ observaciones, y $\alpha = .05$, se asignaría .025 a cada cola de la distribución de t y se rechazaría cuando $t > 2.045$ ó $t < -2.045$. Note que este valor es muy cercano al de $z_{.025} = 1.96$ empleado cuando se usa z como estadística de prueba.

Es importante notar que la t de Student y los valores críticos correspondientes que se han tabulado se basan en la suposición de que la población muestreada posee una distribución de probabilidad normal. Esta resulta ser una suposición muy restrictiva ya que en muchas situaciones reales las propiedades de la población son enteramente desconocidas y puede resultar que la distribución no sea normal. Si esto afectara seriamente la distribución de la estadística

La aplicación de la prueba de t sería muy limitada. Afortunadamente este problema es de escasa consecuencia puesto que puede demostrarse que la distribución de la estadística t posee una forma que resulta muy cercana a la de la distribución teórica de t para poblaciones que si bien no son normales, poseen una distribución de probabilidad de forma acampanada. Esta propiedad de la estadística t y lo frecuente que resulta la ocurrencia de distribuciones acampanadas en datos observados de la naturaleza refuerzan el valor de la t de Student para su uso en la inferencia estadística.

Debe notarse también que $\bar{y}$ y s^2 deben ser independientes (en el sentido probabilístico) para que la cantidad

$$\frac{\bar{y} - \mu}{s/\sqrt{n}}$$

siga una distribución t como distribución muestral. Como se mencionó previamente, este requisito se satisface automáticamente cuando la muestra se toma de una población normal.

Una vez discutido el origen de la distribución t de Student y sus correspondientes valores críticos (tabla 4 en el apéndice), se enfocará de nuevo el problema de hacer inferencias acerca del peso medio de los diamantes en base a una muestra de $n = 6$ observaciones. Antes de considerar la solución, se construirá el equipo conceptual apropiado para hacer inferencia y considerar las seis observaciones, y así llegar a una decisión acerca del significado de los datos.

9.3 Inferencias con muestras pequeñas para la media de una población

La prueba de hipótesis estadística para la media de una población puede establecerse de la manera siguiente:

Prueba con muestras pequeñas para la media de una población

Hipótesis nula: $H_0: \mu = \mu_0$

$$\text{Estadística de prueba: } t = \frac{\bar{y} - \mu_0}{s/\sqrt{n}}$$

Hipótesis alternativa H_a : Determinada por el experimentador dependiendo de los valores alternativos que se deseen detectar.

Región de rechazo: Buscar los valores críticos de t en la tabla 4 del apéndice.

La media y la desviación estándar para los seis diamantes eran .53 y .0559, respectivamente, y los elementos de la prueba definidos en el cuadro de arriba quedan como sigue:

Hipótesis nula $H_0: \mu = .5$

Estadística de prueba:
$$t = \frac{\bar{y} - \mu_0}{s/\sqrt{n}} = \frac{.53 - .5}{.0559/\sqrt{6}} = 1.31.$$

Hipótesis alternativa $H_a: \mu > .5$

Región de rechazo: para $\alpha = .05$ es $t > 2.015$.

Puesto que el valor calculado de la estadística de prueba no cae dentro de la región de rechazo, no se rechaza H_0 . Es decir, los datos no muestran evidencia suficiente para indicar que el peso promedio de los diamantes excede a los 0.5 kilates.

El cálculo de la probabilidad β para el error de tipo II para la prueba de t es difícil y se encuentra fuera de los objetivos de este texto. Por lo tanto se evita este problema y se calcula el estimador por intervalo para μ , como se sugiere en la sección 8.15.

La forma para el intervalo de confianza para μ en el caso de muestras grandes era de la forma

$$\bar{y} \pm z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$

en donde $z_{\alpha/2} = 1.96$ para un coeficiente de confianza (confiabilidad) de .95. Este resultado supone que σ es conocido y sólo requiere una medida de $1.96\sigma_{\bar{y}}$ (ó $2\sigma_{\bar{y}}$ aproximadamente) a cada lado de $\bar{y}$, de acuerdo a la regla empírica. Cuando σ se desconoce y debe estimarse mediante s usando una muestra pequeña, el intervalo basado en z no contendrá a μ 95% de las veces si se reprodujera el proceso de inferencia en repetidas ocasiones. Aunque no se deriva aquí, resultará claro que el intervalo correspondiente a una muestra pequeña es el que se presenta en el siguiente cuadro.

Intervalo de confianza del $(1 - \alpha)$ 100% para μ con base en una muestra pequeña

$$\bar{y} \pm t_{\alpha/2} \frac{s}{\sqrt{n}}$$

donde $s/\sqrt{n}$ es la desviación estándar estimada de $\bar{y}$.

Para nuestro ejemplo el intervalo con 95% de confiabilidad para μ es

$$\bar{y} \pm t_{\alpha/2} \frac{s}{\sqrt{n}} = .53 \pm 2.571 \frac{.0559}{\sqrt{6}} = .53 \pm .059$$

Por lo que el estimador por intervalo para μ es de .471 a .589 con confiabilidad de .95. Si el experimentador desea detectar un pequeño incremento en el peso de los diamantes, incremento en exceso de 0.5 kilates, la longitud del intervalo debe reducirse mediante más observaciones del peso de los diamantes. Con esto disminuirán $1/\sqrt{n}$ y $t_{\alpha/2}$ y por lo tanto disminuirá la longitud del intervalo. O, equivalentemente, desde el punto de vista de la prueba de hipótesis, se tendrá más información en la cual basar la decisión y así la probabilidad de cometer el error de tipo II decrecerá.

Ejemplo 9.1

Un fabricante de pólvora ha desarrollado un nuevo tipo de pólvora diseñado para producir una velocidad de salida del proyectil de 3000 metros por segundo. Se cargan ocho proyectiles con la pólvora y se registran las velocidades de salida. Las velocidades resultantes se muestran en la tabla.

VELOCIDADES DE SALIDA	(METROS/SEGUNDO)
3005	2995
2925	3005
2935	2935
2965	2905

¿Presentan estos datos suficiente evidencia para indicar que la velocidad promedio es diferente de 3000 m/seg?

Solución Se trata de probar la hipótesis nula de que $\mu = 3000$, contra la alternativa de que μ es mayor o menor que 3000 m/seg por lo que se requiere de una prueba de dos colas. Se tiene:

$$H_0: \mu = 3000$$

$$H_a: \mu \neq 3000$$

Usando $\alpha = .05$ y asignando .025 a cada cola de la distribución t se encuentra que el valor crítico de t para $n = 8$ observaciones [o $(n - 1) = 7$ g.l.] es $t = 2.365$. Por lo tanto se rechaza H_0 si $t > 2.365$ o si $t < -2.365$. (Recuerde que la distribución de t es simétrica con respecto a 0.)

La media y la desviación estándar para estos datos son:

$$\bar{y} = 2958.75 \quad \text{y} \quad s = 39.26$$

Entonces

$$t = \frac{\bar{y} - \mu_0}{s/\sqrt{n}} = \frac{2958.75 - 3000}{39.26/\sqrt{8}} = -2.97$$

Puesto que el valor observado para la estadística de prueba se encuentra en la región crítica, se rechaza H_0 y se concluye que la velocidad promedio es menor que 3000 m/seg. Además, se puede estar razonablemente confiado de que se ha tomado una decisión correcta. Al usar este procedimiento sólo se rechazará H_0 equivocadamente el 5% de las veces en repeticiones sucesivas.

El calcular un intervalo de confianza del 95% proporciona información adicional acerca de μ . El intervalo está dado por:

$$\bar{y} \pm t_{\alpha/2} \frac{s}{\sqrt{n}}$$

y substituyendo se tiene

$$2958.75 \pm (2.365) \frac{39.26}{\sqrt{8}} = 2958.75 \pm 32.83$$

Por lo que la velocidad de salida del proyectil se encuentra en el intervalo de 2925.92 a 2991.58 m/seg con .95 de confiabilidad. Se puede obtener un estimador más preciso al aumentar el tamaño de la muestra.

Ejercicios

9.1. ¿Cuándo se considera inadecuado usar z (y la tabla normal) como estadística de prueba en relación al tamaño de muestra? ¿Por qué?

9.2. Las especificaciones para una nueva aleación de alta resistencia al calor establecen que la cantidad de cobre en la aleación debe ser menor del 23.2%. Una muestra de 10 análisis de un lote del producto presenta una media de contenido de cobre de 23.0% y una desviación estándar de .24%. ¿Proporcionan estos análisis evidencia suficiente para indicar que la cantidad de cobre en este lote es menor que el límite especificado? Use $\alpha = .10$.

9.3. Con los datos del ejercicio 9.2 estime el contenido medio de cobre en este lote, usando un intervalo de confianza del 90%.

9.4. Un muestreo aleatorio de $n = 24$ artículos en un supermercado presenta una diferencia entre el valor real del artículo y el valor marcado en éste. La media y la desviación estándar de las diferencias entre el precio real y el marcado en los 24 artículos son $-\$37.14$ y $\$6.42$ respectivamente. Encuentre un intervalo de confianza para la diferencia media entre el valor real y el marcado por artículo en ese supermercado. Use $\alpha = .05$.

9.5. Un contratista ha construido un gran número de casas de aproximadamente el mismo tamaño y el mismo precio. El contratista afirma que el valor promedio de estas casas (o casas similares que él haya construido) no excede de $\$35,000$. Un corredor de bienes raíces selecciona aleatoriamente cinco de las casas construidas recientemente por el contratista y averigua sus precios, obteniendo $\$34,500$, $\$37,000$, $\$36,000$, $\$35,000$, y $\$35,500$. ¿Contradican estas cinco observaciones la afirmación del contratista acerca del valor medio de sus casas? Haga la prueba con $\alpha = .05$ de nivel de significancia.

9.6. Una cadena de estaciones de servicio automotriz trata de implantar un plan de salarios que proporcione un incentivo a sus gerentes y que a la vez resulte competitivo con los de posiciones similares en

las compañías competidoras. Una muestra aleatoria de 12 gerentes de estación para esa compañía arroja un salario medio de $\$16,750$ con desviación estándar de $\$3,100$. ¿Sugieren estos datos que el salario promedio de los gerentes de estación de esa compañía difiere de $\$18,500$, el salario anunciado por la compañía competidora? Pruebe la hipótesis nula de que $\mu = \$18,500$ contra la alternativa de que $\mu \neq \$18,500$ al 5% de nivel de significancia.

9.7. La utilidad por cada auto nuevo vendido por un vendedor varía de auto a auto. La utilidad promedio por venta registrada en la semana pasada fue (en cientos de dólares)

2.1, 3.0, 1.2, 6.2, 4.5, 5.1

Calcule un intervalo de confianza del 90% para la utilidad promedio por venta.

9.8. Siguiendo con el ejercicio 9.7, suponga que la utilidad promedio por venta que desea alcanzar el vendedor es de $\$480$. Con base en la información disponible, ¿hay suficiente evidencia para indicar que el vendedor no ha alcanzado su utilidad objetivo? Use el 5% de significancia.

9.9. Un nuevo aditivo para la gasolina ha sido desarrollado por una compañía norteamericana. Se afirma que el aditivo resulta en por lo menos un 15% de ahorro en gasolina.* En un experimento de uso del aditivo realizado en 8 autos durante un período de una semana se registraron los siguientes porcentajes de ahorros en el consumo de gasolina

15.2	14.1	13.7	15.2
18.6	15.0	14.5	13.8

¿Contradican estos datos la afirmación del fabricante? Use $\alpha = .05$.

9.10. Use un intervalo de confianza del 95% para estimar el ahorro promedio de gasolina en el ejercicio 9.9.

*"Oregonian," 30 de octubre de 1975.

9.4 Inferencias con muestras pequeñas para la diferencia entre dos medias

El contexto del problema que se considera aquí es idéntico al que se discutió en la sección 8.9. Se consideran muestras aleatorias independientes de n_1 y n_2

observaciones respectivamente tomadas de dos poblaciones con medias y varianzas μ_1, σ_1^2 y μ_2, σ_2^2 . El objetivo es hacer inferencias acerca de la diferencia $(\mu_1 - \mu_2)$ entre las medias de ambas poblaciones.

Los métodos para muestras pequeñas de prueba de hipótesis e intervalos de confianza para la diferencia de dos medias están, como en el caso de una sola media, basados en suposiciones acerca de las distribuciones de probabilidad de las poblaciones muestreadas. Específicamente, **se supone aquí que ambas poblaciones poseen una distribución de probabilidad normal y también que estas poblaciones poseen varianzas iguales $\sigma_1^2 = \sigma_2^2$** . Es decir, se supone que la variabilidad en ambas poblaciones es la misma y puede ser representada por una varianza común, a la que se designará como σ^2 . Esto es, $\sigma^2 = \sigma_1^2 = \sigma_2^2$. Aunque esta suposición de igualdad de varianzas puede parecer restrictiva, resulta bastante razonable en muchas situaciones prácticas.

El estimador puntual de $(\mu_1 - \mu_2)$, $(\bar{y}_1 - \bar{y}_2)$, la diferencia entre las medias muestrales, se discutió en la sección 8.9 y se observó ahí que resultaba ser insesgado y poseía una desviación estándar de

$$\sigma_{(\bar{y}_1 - \bar{y}_2)} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

para su distribución muestral. Este resultado se utilizó para determinar la cota para el error de estimación, para la construcción de un intervalo de confianza para muestras grandes y para la estadística de prueba z ,

$$z = \frac{(\bar{y}_1 - \bar{y}_2) - D_0}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

para probar la hipótesis $H_0: \mu_1 - \mu_2 = D_0$, en donde D_0 es la diferencia, dada por la hipótesis, entre las medias. Usando la suposición de que $\sigma_1^2 = \sigma_2^2 = \sigma^2$, la estadística de prueba se puede simplificar como sigue:

$$z = \frac{(\bar{y}_1 - \bar{y}_2) - D_0}{\sqrt{\frac{\sigma^2}{n_1} + \frac{\sigma^2}{n_2}}} = \frac{(\bar{y}_1 - \bar{y}_2) - D_0}{\sigma \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

Resulta razonable basar la prueba de la hipótesis (con muestras pequeñas) $H_0: \mu_1 - \mu_2 = D_0$ en la estadística que se presenta en el siguiente cuadro.

Estadística de prueba t para muestras pequeñas

$$t = \frac{(\bar{y}_1 - \bar{y}_2) - D_0}{s \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

Esto es, se substituye el estimador s en lugar de σ . Sorprendentemente esta estadística de prueba posee también una distribución muestral t de Student cuando se satisfacen las suposiciones arriba establecidas. Este hecho puede ser

demostrado matemáticamente o verificado experimentalmente al muestrear dos poblaciones normales.

El estimador s que se usa en la estadística de prueba podría ser ya sea s_1 o s_2 , la desviación estándar de cualquiera de las muestras, aunque el uso de cualquiera de estos sería un desperdicio puesto que ambos estiman a σ . Puesto que se desea obtener el mejor estimador disponible resulta razonable usar un estimador que combine la información obtenida en ambas muestras. En el siguiente cuadro se muestra este estimador que utiliza la suma de los cuadrados de las desviaciones alrededor de la media para ambas muestras.

Estimador combinado para σ^2

$$s^2 = \frac{\sum_{i=1}^{n_1} (y_i - \bar{y}_1)^2 + \sum_{i=1}^{n_2} (y_i - \bar{y}_2)^2}{n_1 + n_2 - 2}$$

en donde

$$s_1^2 = \frac{\sum_{i=1}^{n_1} (y_i - \bar{y}_1)^2}{n_1 - 1} \quad y \quad s_2^2 = \frac{\sum_{i=1}^{n_2} (y_i - \bar{y}_2)^2}{n_2 - 1}$$

Note que este estimador combinado puede escribirse como

$$s^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$$

Como en el caso de una sola muestra, el denominador $(n_1 + n_2 - 2)$ en la fórmula de s^2 se denomina “número de grados de libertad” asociado con s^2 . Se puede demostrar matemáticamente que el valor esperado del estimador combinado s^2 es igual a σ^2 y por eso s^2 es un estimador insesgado de la varianza común de las poblaciones. Por último debe notarse que los divisores de las sumas de cuadrados de las desviaciones en s_1^2 y s_2^2 , $(n_1 - 1)$ y $(n_2 - 1)$ respectivamente, son los números de grados de libertad asociados con estos dos estimadores independientes de σ^2 ; es interesante notar que el estimador s que combina la información en ambas muestras posee $(n_1 - 1) + (n_2 - 1) = (n_1 + n_2 - 2)$ grados de libertad.

Prueba estadística de hipótesis, con muestras pequeñas, para la diferencia entre dos medias

$$H_0: \mu_1 - \mu_2 = D_0.$$

$$\text{Estadística de prueba: } t = \frac{(\bar{y}_1 - \bar{y}_2) - D_0}{s \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

La hipótesis alternativa H_a y α son especificados por el experimentador y se usan para localizar los valores críticos de t para la región de rechazo. s es la raíz cuadrada del estimador combinado de σ^2 .

El valor crítico de t puede obtenerse de la tabla 4 del apéndice. Entonces si $n_1 = 10$ y $n_2 = 12$ se usaría el valor de t correspondiente a $(n_1 + n_2 - 2) = 20$ grados de libertad (g.l.). El siguiente ejemplo ilustra este procedimiento de prueba.

Ejemplo 9.2

En los Estados Unidos algunas organizaciones de fabricantes incurren en costos considerables para el entrenamiento de nuevos empleados. No solamente hay un costo directo involucrado en el programa de entrenamiento, sino que, existe también un costo indirecto para la empresa manufacturera puesto que los empleados en entrenamiento no contribuyen directamente en el proceso de manufactura de la empresa. De aquí que estas organizaciones estén en busca de programas que pueden llevar a los empleados a un grado máximo de eficiencia en el menor tiempo posible.

Tabla 9.2 Tiempo (en minutos) de ensamblado de una componente

PROCEDIMIENTO ESTÁNDAR		NUEVO PROCEDIMIENTO	
32	44	35	40
37	35	31	27
35	31	29	32
28	34	25	31
41		34	

Una operación de ensamblado en una planta requiere de un período de entrenamiento de aproximadamente un mes para que un empleado nuevo alcance un grado máximo de eficiencia. Se ha sugerido un nuevo método de entrenamiento y se ha realizado una prueba para comparar el nuevo método con el procedimiento estándar. Con este fin se han entrenado dos grupos de nueve empleados cada uno, durante tres semanas, un grupo siguiendo el nuevo método y el otro con el procedimiento estándar. El tiempo (en minutos) de armado de la componente se registró para cada empleado al final del período de tres semanas. Estas observaciones se muestran en la tabla 9.2. ¿Presentan estos datos evidencia suficiente de que el tiempo medio de ensamblado es menor para el nuevo procedimiento de entrenamiento?

Solución Sean μ_1 y μ_2 el tiempo promedio de ensamblado para el procedimiento estándar y para el nuevo respectivamente. Entonces puesto que se busca evidencia en favor de que $\mu_1 > \mu_2$, se probará la hipótesis nula $H_0: \mu_1 = \mu_2$ (i.e. $\mu_1 - \mu_2 = 0$) contra la hipótesis alternativa $H_a: \mu_1 > \mu_2$ (i.e. $\mu_1 - \mu_2 > 0$). Para realizar la prueba suponga que la variabilidad en el tiempo medio de ensamblado es en esencia una función de las diferencias entre individuos (no debidas a los programas), que las poblaciones de observaciones tienen distribución aproximadamente normal y que la variabilidad para las dos poblaciones de observaciones es aproximadamente igual.

Las medias muestrales y las sumas de cuadrados de las observaciones son

$$\bar{y}_1 = 35.22 \quad \text{y} \quad \sum_{i=1}^9 (y_i - \bar{y}_1)^2 = 195.56$$

$$\bar{y}_2 = 31.56 \quad y \quad \sum_{i=1}^9 (y_i - \bar{y}_2)^2 = 160.22$$

Entonces el estimador combinado para la varianza resulta

$$s^2 = \frac{\sum_{i=1}^9 (y_i - \bar{y}_1)^2 + \sum_{i=1}^9 (y_i - \bar{y}_2)^2}{n_1 + n_2 - 2} = \frac{195.56 + 160.22}{9 + 9 - 2} = 22.24$$

y la desviación estándar es $s \approx 4.72$.

La hipótesis alternativa $H_a: \mu_1 > \mu_2$, o equivalentemente $\mu_1 - \mu_2 > 0$ implica que se debe usar una prueba estadística de una sola cola y que la región de rechazo para la prueba se localiza en la cola superior de la distribución de t . Buscando en la tabla 4 del apéndice se obtiene el valor crítico de t para $\alpha = .05$ y $(n_1 + n_2 - 2) = 16$ grados de libertad. Este resulta 1.746. Por lo tanto se rechaza la hipótesis nula cuando el valor calculado de t sea mayor que 1.746.

El valor calculado de la estadística de prueba es

$$t = \frac{(\bar{y}_1 - \bar{y}_2)}{s \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = \frac{35.22 - 31.56}{4.72 \sqrt{\frac{1}{9} + \frac{1}{9}}} \approx 1.64$$

Comparando éste con el valor crítico $t_{.05} = 1.746$, se nota que el valor calculado no cae en la región de rechazo, por lo que se debe concluir que la evidencia resulta insuficiente para indicar que el nuevo método de entrenamiento es superior, al nivel de .05 de significancia.

El intervalo de confianza en base a muestras pequeñas para $(\mu_1 - \mu_2)$ se basa en las mismas suposiciones que el procedimiento de prueba estadística. El intervalo de confianza con coeficiente de confiabilidad de $(1 - \alpha)$ se presenta en el siguiente cuadro.

Intervalo de confianza del $(1 - \alpha)100\%$ con base en muestras pequeñas para $(\mu_1 - \mu_2)$

$$(\bar{y}_1 - \bar{y}_2) \pm t_{\alpha/2} s \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

Nótese la similitud en los procedimientos para construir el intervalo de confianza para una sola media, sección 9.3, y para la diferencia entre dos medias. En ambos casos el intervalo se construye usando el estimador puntual apropiado y sumando y restando una cantidad de $t_{\alpha/2}$ veces la estimación de la desviación estándar del estimador.

Ejemplo 9.3

Haciendo referencia al ejemplo 9.2, puesto que existen un gasto y un riesgo asociados al desechar un procedimiento de entrenamiento, sería conveniente

para los administradores estimar la diferencia entre las dos medias en tiempo de ensamblado por medio de un intervalo de confianza de .95 de confiabilidad. Encuentre tal intervalo.

Solución Substituyendo en la fórmula

$$(\bar{y}_1 - \bar{y}_2) \pm t_{\alpha/2} s \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

se tiene que el estimador por intervalo (con confiabilidad de .95) es

$$(35.22 - 31.56) \pm (2.120)(4.72) \sqrt{\frac{1}{9} + \frac{1}{9}} = 3.66 \pm 4.71$$

Entonces se estima que la diferencia $(\mu_1 - \mu_2)$ en tiempos de ensamblado cae en el intervalo de -1.05 a 8.37 . Debe notarse que la longitud del intervalo es considerable y parece aconsejable aumentar el tamaño de muestra y reestimar el intervalo.

Antes de concluir esta discusión es necesario comentar acerca de las dos suposiciones en las que se basan estos procedimientos de inferencia. **Cuando se viola moderadamente la suposición de que las poblaciones poseen una distribución de probabilidad normal no se afecta seriamente la distribución de la estadística de prueba ni la confiabilidad para el correspondiente intervalo de confianza. En cambio las varianzas de las poblaciones deben ser casi iguales para asegurar la validez de los procedimientos anteriores.**

Si hubiese alguna razón para creer que las varianzas de las poblaciones son distintas, se debe proceder a ajustar los procedimientos de prueba y los intervalos de confianza correspondientes. La discusión de estas técnicas se omite aquí y se refiere al lector al texto de Li (1961) que se lista en las referencias.

En la sección 9.7 se presentará un procedimiento para probar hipótesis acerca de la igualdad de las varianzas de dos poblaciones.

Ejercicios

9.11. Una planta manufacturera tiene dos máquinas ensambladoras que realizan operaciones idénticas en distintas líneas de ensamblado. Las interrupciones ocurren frecuentemente como resultado del constante uso. Se registraron los tiempos entre 10 descomposturas sucesivas para cada máquina. Suponga que el tiempo entre descomposturas para cada máquina sigue una distribución normal con varianza común σ^2 . En la tabla se muestran las medias y las varianzas de los tiempos entre descomposturas para cada máquina.

¿Presentan estos datos evidencia suficiente para indicar una diferencia en los tiempos medios entre descomposturas? Use $\alpha = .10$.

MAQUINA 1	MAQUINA 2
$\bar{y}_1 = 60.4 \text{ min}$	$\bar{y}_2 = 65.3 \text{ min}$
$s_1^2 = 31.40 \text{ min}^2$	$s_2^2 = 44.82 \text{ min}^2$

9.12. Se hizo una entrevista a cinco subdirectores y a cuatro analistas de mercado de una gran empresa. Se les preguntó a cada uno cuál consideraba ser el porcentaje óptimo de cobertura de mercado para su compañía. Se obtuvieron las siguientes respuestas:

<i>Sub-directores</i>	22.5	25.0	30.0	27.5	20.0
<i>Analistas de mercado</i>	22.5	17.5	17.0	20.0	

¿Sugieren estos datos que los subdirectores y los analistas de mercado están en desacuerdo cuando estiman la cobertura óptima de mercado para la empresa? Pruebe al 5% de significancia.

9.13. La productividad en el trabajo depende fuertemente de muchos y muy variados factores, tales como el salario, la complejidad de la operación y el ambiente en el trabajo. Pero es a menudo el diseño de la operación (i.e., la secuencia ordenada de movimientos del trabajador y de utilización de material), el factor más importante en la productividad. Dos diseños de operación se someten a consideración para ser implantados en una fábrica. De un estudio de tiempos y movimientos se tiene que de 12 trabajadores usando el diseño A, se tiene una media de 304 segundos y una desviación estándar de 18 segundos y de 15 trabajadores usando el diseño B se tiene una media de 335 segundos con desviación estándar de 24 segundos. ¿Presentan estos datos evidencia suficiente de una diferencia en la tasa de productividad para los dos diseños? Use $\alpha = .01$.

9.14. Refiérase al ejercicio 9.13. Suponga que la planta manufacturera usa actualmente el diseño B para sus operaciones de ensamblado, que el diseño A es experimental y debe considerarse como sustituto del diseño B si se demuestra que es efectivo en cuanto reduce el tiempo de ensamblado. ¿Cambia esta información adicional el tipo de prueba estadística a usar? ¿Cómo?

9.15. El advenimiento de materiales sintéticos tales como el nylon, poliéster y latex, y su introducción en el mercado ha suscitado debates acerca de la cali-

dad y resistencia de estas fibras comparadas con las fibras naturales. Un fabricante de una nueva fibra sintética afirma que su producto posee una mayor resistencia a la tracción que las fibras naturales. Se seleccionaron al azar 10 fibras sintéticas y 10 fibras naturales para determinar su resistencia. Las medias y las varianzas para las dos muestras se presentan en

FIBRA NATURAL	FIBRA SINTÉTICA
$\bar{y}_1 = 272 \text{ kg}$	$\bar{y}_2 = 335 \text{ kg}$
$s_1^2 = 1636 \text{ kg}^2$	$s_2^2 = 1892 \text{ kg}^2$

la tabla. ¿Confirman estos datos la afirmación del fabricante? Haga la prueba usando $\alpha = .10$.

9.16. Refiérase al ejercicio 9.15. Encuentre un intervalo de confianza del 95% para la diferencia en resistencia media entre la fibra natural y la fibra sintética.

9.17. Continuando con los ejercicios 9.15 y 9.16 suponga que se desea estimar la diferencia en resistencia media correcta dentro de 10 kg, con una probabilidad aproximada de .95.

a. Suponiendo que se desean tamaños de muestra iguales para la fibra natural y la fibra sintética, ¿qué tamaño de muestra se requiere para cada fibra?

b. Realizar el experimento con el tamaño de muestra de la parte (a) requiere de gastos adicionales de tiempo y dinero. ¿Es posible reducir los tamaños de muestra especificados en la parte (a) y seguir alcanzando la cota de 10 kg. en el error de estimación?

9.5 Una prueba para diferencias apareadas

Un fabricante desea comparar el rodaje de dos distintos tipos de neumáticos, A y B. Para realizar la comparación se seleccionan aleatoriamente un neumático del tipo A y uno del tipo B y se montan en las ruedas traseras de cada uno de 5 automóviles. Los automóviles se usan por una distancia preespecificada y se registra la cantidad desgastada para cada neumático. Los datos se muestran en la tabla 9.3. ¿Presentan estos datos suficiente evidencia indicando una diferencia en el desgaste promedio de los dos tipos de neumático?

Al analizar los datos se nota que la diferencia entre las dos medias muestrales es de $(\bar{y}_1 - \bar{y}_2) = .48$, cantidad pequeña considerando la variabilidad de los datos y el pequeño número de observaciones consideradas. A primera vista puede parecer que no se tiene evidencia que muestre una diferencia entre las medias de las poblaciones, conjetura que puede verificarse mediante el método presentado en la sección 9.4.

Tabla 9.3 *Desgaste para los dos tipos de neumático*

AUTOMOVIL	TIPO DE NEUMÁTICO	
	A	B
1	10.6	10.2
2	9.8	9.4
3	12.3	11.8
4	9.7	9.1
5	8.8	8.3
	$\bar{y}_1 = 10.24$	$\bar{y}_2 = 9.76$

El estimador combinado de la varianza común s^2 es

$$s^2 = \frac{\sum_{i=1}^{n_1} (y_i - \bar{y}_1)^2 + \sum_{i=1}^{n_2} (y_i - \bar{y}_1)^2}{n_1 + n_2 - 2} = \frac{6.932 + 7.052}{5 + 5 - 2} = 1.748$$

y entonces

$$s \approx 1.32$$

El valor calculado de t que se usa para probar la hipótesis $\mu_1 = \mu_2$ es

$$t = \frac{\bar{y}_1 - \bar{y}_2}{s \sqrt{\frac{1}{n} + \frac{1}{n}}} = \frac{10.24 - 9.76}{1.32 \sqrt{\frac{1}{5} + \frac{1}{5}}} = .58$$

valor que no es ni siquiera cercano a la región de rechazo para la hipótesis $\mu_1 = \mu_2$.

El intervalo de confianza del 95% correspondiente es

$$(\bar{y}_1 - \bar{y}_2) \pm t_{\alpha/2} s \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} = (10.24 - 9.76) \pm (2.306)(1.32) \sqrt{\frac{1}{5} + \frac{1}{5}}$$

ó de -1.45 a 2.41 . Note que la longitud del intervalo es bastante grande considerando la pequeña diferencia entre las medias muestrales.

Un análisis más detenido de los datos revela una inconsistencia con la conclusión anterior. Debe notarse que en cada caso el desgaste medido para el neumático A es mayor que el correspondiente valor para el tipo B para los 5 automóviles. Estas diferencias, denotadas $d = A - B$, se presentan en la tabla 9.4.

Tabla 9.4 *Diferencias en desgaste usando la tabla 9.3*

AUTOMOVIL	$d = A - B$
1	.4
2	.4
3	.5
4	.6
5	.5
	$\bar{d} = .48$

Si se usa como estadística de prueba y , el número de autos para los cuales el desgaste del neumático A es mayor que el del neumático B, como se hizo en el ejercicio 6.30 del capítulo 6, entonces la probabilidad de que A sea mayor que B para un automóvil dado, suponiendo que no hay diferencia en el desgaste de los neumáticos, es $p = \frac{1}{2}$ y entonces y sería una variable aleatoria binomial. Si se escoge $y = 0$ y $y = 5$ como región de rechazo para una prueba de dos colas, entonces $\alpha = p(0) + p(5) = 2(\frac{1}{2})^5 = \frac{1}{16}$. Se rechazará $H_0: \mu_1 = \mu_2$ con una probabilidad de error tipo I de $\alpha = \frac{1}{16}$. Esta es buena evidencia de que existe una diferencia en el desgaste promedio de los dos tipos de neumáticos.

Nótese que se han usado aquí dos estadísticas de prueba distintas para probar la misma hipótesis. ¿No parece extraño que la prueba de t , que utiliza más información (las observaciones de la muestra) que la prueba binomial, no proporciona evidencia suficiente para rechazar la hipótesis nula $\mu_1 = \mu_2$?

La explicación de esta aparente inconsistencia es sencilla. La prueba de t descrita en la sección 9.4 no es la prueba estadística apropiada para este ejemplo. El procedimiento de la sección 9.4 requiere que las muestras sean *aleatorias e independientes*. Aquí la suposición de independencia fue violada por la manera en que se realizó el experimento, puesto que el par de observaciones una para A y otra para B en cada automóvil están seguramente relacionadas. Al observar los datos puede notarse que las lecturas tienen aproximadamente la misma magnitud para cada automóvil pero varían de un automóvil a otro. Esto es por supuesto lo que podría esperarse, ya que el desgaste de los neumáticos depende en gran parte de los hábitos al conducir, el balanceo de las ruedas y la superficie sobre la cual se llevó a cabo la prueba. Puesto que cada automóvil tiene distinto piloto, es de esperarse una gran variabilidad de un automóvil a otro.

La experiencia ganada con la estimación por intervalo ha mostrado que la longitud de los intervalos depende de la magnitud de la desviación estándar del estimador puntual del parámetro. Entre más pequeña esta magnitud mejor la estimación y más factible que la estadística de prueba proporcione evidencia para rechazar la hipótesis nula, si ésta es, en verdad, falsa. El conocimiento de este fenómeno fue utilizado para *diseñar* el experimento de desgaste de los neumáticos. El experimentador advirtió que las observaciones de dicho desgaste variarían grandemente de un auto al otro y que esta variabilidad no puede ser considerada en los datos si los neumáticos se asignaran a las 10 ruedas de manera aleatoria (si este hubiera sido el caso, el análisis de acuerdo al procedimiento de la sección 9.4 hubiera sido apropiado). En cambio, una comparación del desgaste entre los neumáticos A y B puede hacerse para cada automóvil resultando cinco diferencias de observaciones. El diseño elimina el efecto de la variabilidad de auto a auto y produce mayor información acerca de la diferencia media del desgaste para los dos tipos de neumáticos.

Un análisis apropiado de los datos utilizaría las cinco diferencias en las observaciones para probar la hipótesis de que la diferencia media es igual a cero, proposición que resulta equivalente a la hipótesis nula $\mu_1 = \mu_2$.

Como puede verificarse, el promedio y la desviación estándar de las 5 diferencias observadas son :

$$\bar{d} = .48 \quad \text{y} \quad s_d = .0837$$

Entonces

$$H_0: \mu_d = 0$$

y

$$t = \frac{\bar{d} - 0}{s_d / \sqrt{n}} = \frac{.48}{.0837 / \sqrt{5}} = 12.8$$

El valor crítico de t para una prueba estadística con $\alpha = .05$ y 4 grados de libertad es 2.776. Ciertamente el valor de t es bastante grande y altamente significativo, por lo que se concluye que el desgaste medio del neumático B es menor que el del tipo A.

Un intervalo de confianza del 95% para la diferencia en el desgaste medio es

$$\bar{d} \pm t_{\alpha/2} \frac{s_d}{\sqrt{n}} = .48 \pm (2.776) \frac{.0837}{\sqrt{5}} = .48 \pm .10$$

El diseño estadístico del experimento de los neumáticos es un ejemplo simple de un **diseño aleatorizado en bloques** y la estadística de prueba resultante se denomina a menudo **prueba de diferencias apareadas**. Debe notarse que el **apareamiento de los datos ocurrió al planearse el experimento y no después de que los datos fueron colectados**. Las comparaciones del desgaste de los neumáticos se hicieron dentro de bloques relativamente homogéneos (automóviles) y cada tipo de neumático se asignó al azar a cada una de las ruedas traseras del automóvil.

Como una indicación de la ganancia en cantidad de información obtenida al introducir los bloques en el experimento puede compararse el intervalo de confianza obtenido para el análisis no apareado (incorrecto), con el intervalo obtenido para las diferencias apareadas. El intervalo de confianza para $(\mu_1 - \mu_2)$ que se obtendría al haber asignado aleatoriamente los neumáticos a las 10 ruedas (no apareado) es desconocido, pero probablemente tendría una longitud similar a la del intervalo -1.45 a 2.41 , que se obtuvo al analizar los datos en forma no apareada. El apareamiento de los tipos de neumáticos en los automóviles (bloques) y el análisis resultante de las diferencias produjo el intervalo de confianza de $.38$ a $.58$. Resulta notable la diferencia en longitud de ambos intervalos, lo que indica el muy considerable aumento en información obtenido al considerar los bloques en el experimento.

Aunque en el experimento anterior el introducir los bloques resultó en beneficio considerable, éste no siempre es el caso. Debe observarse que los grados de libertad para la estimación de σ^2 son menos para las diferencias apareadas que para el experimento no apareado. Si realmente no existiese diferencia entre los bloques, la reducción en grados de libertad produciría un aumento en el valor de $t_{\alpha/2}$ empleado en el intervalo de confianza aumentando así la longitud del intervalo. Esto, por supuesto, no ocurrió en el experimento anterior puesto que la reducción en la desviación estándar de $\bar{d}$ compensó por mucho la pérdida en grados de libertad.

Se concluye haciendo énfasis en la siguiente observación. Una vez que se ha usado un diseño apareado para un experimento ya no se tiene la opción de usar el análisis no apareado de la sección 9.4, puesto que las suposiciones en las que se basa esa prueba (muestras independientes) han resultado violadas. La única alternativa es usar el método correcto de análisis, diferencias apareadas, (y el correspondiente intervalo de confianza) descrito en esta sección.

Ejercicios

9.18. Para comparar la efectividad de un programa de seguridad en el trabajo, se observó en seis distintas plantas el número de accidentes por mes antes y después del programa. Los datos aparecen en la tabla siguiente. ¿Proporcionan estos datos evidencia suficiente para indicar que el programa ha sido efectivo al reducir el número de accidentes laborales por mes? Haga la prueba al $\alpha = .10$ de nivel de significancia.

9.19. Considere la prueba de hipótesis del ejercicio 9.18 desde el punto de vista de (i) el empresario,

quien desea estar absolutamente seguro de los efectos benéficos del programa para justificar el costo de su implementación, y (ii) del gobierno, que desea proporcionar a los trabajadores suficiente seguridad para su salud a cualquier costo razonable. Recomiende la selección de un nivel α de significancia para la prueba de hipótesis acerca de la efectividad del programa para ambos puntos de vista:

a. el empresario b. el gobierno

	PLANTA NUMERO					
	1	2	3	4	5	6
<i>Antes del programa</i>	38	64	42	70	58	30
<i>Después de programa</i>	31	58	43	65	52	29

9.20. Se desea determinar el contenido de grasa en la carne para poder determinar su precio de venta al consumidor. Una compañía empacadora de carne está considerando el uso de dos métodos diferentes para determinar el porcentaje de grasa. Ambos métodos fueron usados para evaluar el contenido de grasa en ocho diferentes muestras de carne. Los resultados se muestran en la siguiente tabla. ¿Sugieren estos

datos que los dos métodos difieren en su medición del contenido de grasa en la carne? Pruebe al $\alpha = .05$ de nivel de significancia.

9.21. Continuando con el ejercicio 9.20 use un intervalo de 90% de confianza para estimar la diferencia en contenido de grasa medio obtenido al usar los dos métodos.

	MUESTRA							
PORCENTAJE DE GRASA USANDO	1	2	3	4	5	6	7	8
Metodo 1	23.1	27.1	25.0	27.6	22.2	27.1	23.2	24.7
Metodo 2	22.7	27.4	24.9	27.2	22.5	27.4	23.6	24.4

9.22. Se seleccionan al azar cinco, entre todas las secretarías de una compañía de seguros. Se registra la velocidad en la mecanografía (en palabras por minuto), para cada secretaria en una máquina eléctrica y en una máquina estándar. Los datos se presentan en la siguiente tabla. ¿Justifican estos datos la conclusión de que la velocidad de mecanografiar de las secretarías es mayor al usar la máquina eléctrica que al usar la máquina estándar? Use $\alpha = .10$.

9.23. ¿Por qué se usan observaciones apareadas para estimar la diferencia entre las medias de dos poblaciones en lugar de usar la estimación basada en muestras independientes seleccionadas de ambas poblaciones? Por ejemplo, en el ejercicio 9.22 ¿cuál es la ventaja de usar las mismas cinco secretarías para ambos factores del experimento en lugar de usar 5 diferentes secretarías para cada factor? ¿Resulta siempre preferible un experimento apareado? Explique.

SECRETARIA	MÁQUINA ELÉCTRICA	MÁQUINA ESTÁNDAR
1	82	73
2	77	69
3	79	75
4	68	62
5	84	71

9.6 Inferencias acerca de la varianza de una población

Se ha visto en secciones anteriores que el estimar la varianza σ^2 resulta fundamental para los procedimientos de inferencia acerca de las medias de las poblaciones. Más aún, existen muchas situaciones prácticas en donde σ^2 es el objetivo primario de la investigación experimental. En estos casos σ^2 adquiere una posición de mayor importancia que la media de la población.

Los instrumentos científicos de medición deben proporcionar lecturas insesgadas con un muy pequeño error de medición. El altímetro de un aeroplano que, en promedio, mide la altitud correctamente sería de poca utilidad si la desviación estándar del error de medición fuese de 2000 metros. Tal vez un sesgo en la medición puede ser corregida al calibrar el instrumento, pero la precisión del instrumento medido por la desviación estándar del error de medición, está usualmente en función del diseño del instrumento mismo y no puede ser controlada.

Las partes producidas por un proceso de manufactura deben ser producidas con un mínimo de variabilidad para reducir el número de productos fuera del rango aceptable (defectuosos). En general se desea mantener una varianza mínima en las mediciones de las características de calidad de un producto industrial para alcanzar el control del proceso y minimizar el porcentaje de productos de baja calidad.

La varianza muestral

$$s^2 = \frac{\sum_{i=1}^n (y_i - \bar{y})^2}{n - 1}$$

es un estimador insesgado de la varianza de la población σ^2 . La distribución muestral de s^2 , generada mediante repetidas muestras, es una distribución de probabilidad que empieza en $s^2 = 0$ (ya que s^2 no puede ser negativa) con media igual a σ^2 . A diferencia de la distribución de $\bar{y}$, la distribución de s^2 no es simétrica. La forma exacta de la distribución depende de la distribución de probabilidad de la población de la cual las observaciones en la muestra fueron tomadas.

En la metodología que se describe a continuación, se supone que la muestra fue tomada de una población normal y que s^2 está calculada con base en una muestra aleatoria de n observaciones. Usando la terminología de la sección 9.2 se dirá que s^2 tiene asociados $(n - 1)$ grados de libertad.

El siguiente paso es considerar la distribución muestral de s^2 , en muestreo repetido, para una distribución normal dada (con una media y una varianza fijas), y tabular valores de s^2 para algunas de las más usuales áreas bajo la curva. Al hacer esto se encontrará que la distribución de s^2 es independiente de la media de la población μ , pero posee una distribución diferente para cada tamaño de muestra y para cada valor de σ^2 . Esta labor requiere de mucho trabajo pero afortunadamente puede simplificarse mediante una **estandarización** como se hizo al usar z en las tablas normales. La cantidad usada para esta estandarización se presenta en el siguiente cuadro.

Una variable aleatoria ji-cuadrada

$$\chi^2 = \frac{(n - 1)s^2}{\sigma^2}$$

Tabla 9.5 Formato de la tabla de ji-cuadrada, tabla 5 del apéndice

g.l.	$\chi^2_{0.995}$	$\chi^2_{0.950}$	$\chi^2_{0.900}$	$\chi^2_{0.100}$	$\chi^2_{0.050}$	$\chi^2_{0.005}$	g.l.
1	0.0000393	0.0039321	0.0157908	2.70554	3.84146	7.87944	1
2	0.0100251	0.102587	0.210720	4.60517	5.99147	10.5966	2
3	0.0717212	0.351846	0.584375	6.25139	7.81473	12.8381	3
4	0.206990	0.710721	1.063623	7.77944	9.48773	14.8602	4
5	0.411740	1.145476	1.61031	9.23635	11.0705	16.7496	5
6	0.675727	1.63539	2.20413	10.6446	12.5916	18.5476	6
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
15	4.60094	7.26094	8.54675	22.3072	24.9958	32.8013	15
16	5.14224	7.96164	9.31223	23.5418	26.2962	34.2672	16
17	5.69724	8.67176	10.0852	24.7690	27.5871	35.7185	17
18	6.26481	9.39046	10.8649	25.9894	28.8693	37.1564	18
19	6.84398	10.1170	11.6509	27.2036	30.1435	38.5822	19

La cantidad χ^2 , a la que los estadísticos llaman una **variable ji-cuadrada** (χ es la letra griega ji) se ajusta admirablemente a nuestros propósitos. Su distribución en muestreo repetido es llamada, como podría ya sospecharse, la **distribución ji-cuadrada**. La ecuación de esta distribución muestral es ampliamente conocida por los estadísticos, quienes han tabulado los valores correspondientes a las áreas de interés bajo la curva de la distribución. Estos valores se presentan en la tabla 5 del apéndice. El formato de la tabla 5 del apéndice se encuentra en la tabla 9.5.

La forma de la distribución ji-cuadrada, como ocurre con la distribución de t , varía con el tamaño de muestra, o equivalentemente con los grados de libertad asociados a s^2 . Entonces, la tabla 5 se construye en la misma forma que la tabla de t , con los grados de libertad en la primera y última columnas. El símbolo χ^2_α indica que el valor- χ^2 tabulado es tal que una área α queda a su derecha (ver figura 9.3). Hablando en términos probabilísticos,

$$P(\chi^2 > \chi^2_\alpha) = \alpha$$

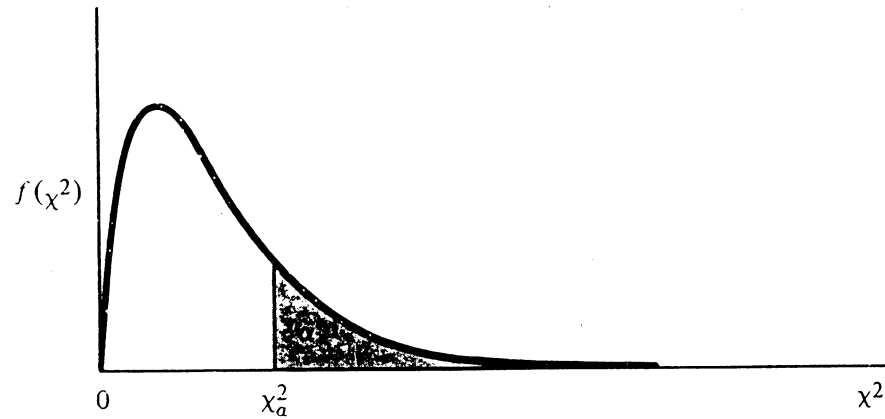


Figura 9.3 Una distribución ji-cuadrada

Entonces 99% del área bajo la curva de la distribución χ^2 queda a la derecha de $\chi^2_{.99}$. Nótese que los valores extremos de la distribución de χ^2 deben tabularse para ambas, la cola inferior y la cola superior, puesto que la distribución no es simétrica.

Puede usted comprobar su habilidad al verificar las siguientes proposiciones. La probabilidad de que χ^2 basada en $n = 16$ observaciones (15 g.l.), exceda 24.9958 es .05. Para una muestra de $n = 6$ observaciones (5 g.l.), 95% del área bajo la distribución χ^2 queda a la derecha de $\chi^2 = 1.145476$. Estas cifras se encuentran marcadas en la tabla 9.5.

La prueba estadística de una hipótesis nula acerca de la varianza de una población

$$H_0: \sigma^2 = \sigma_0^2$$

usa como estadística de prueba

$$\chi^2 = \frac{(n-1)s^2}{\sigma_0^2}$$

Si σ^2 es realmente mayor que el valor dado por la hipótesis, σ_0^2 , la estadística tenderá a tomar valores grandes y probablemente caerá en la cola superior de la distribución. Si $\sigma^2 < \sigma_0^2$ la estadística de prueba tenderá a ser pequeña y probablemente caerá hacia la cola inferior de la distribución. Como en otras pruebas estadísticas se puede usar una prueba de una o dos colas dependiendo de la hipótesis alternativa que se escoja.

Prueba de una hipótesis acerca de la varianza de una población

Hipótesis nula $H_0: \sigma^2 = \sigma_0^2$

Hipótesis alternativa $H_a: \sigma^2 \neq \sigma_0^2$, para una prueba de dos colas.

$$\text{Estadística de prueba: } \chi^2 = \frac{(n-1)s^2}{\sigma_0^2}$$

Región de rechazo: Para una prueba de dos colas, rechace si $\chi^2 < \chi^2_{1-\alpha/2}$, o si $\chi^2 > \chi^2_{\alpha/2}$ en donde $\chi^2_{1-\alpha/2}$ es el valor de χ^2 tal que $100(1 - \alpha/2)\%$ del área de la distribución χ^2 con $n - 1$ grados de libertad queda a la derecha.

Para alternativas $H_a: \sigma^2 > \sigma_0^2$ ó $H_a: \sigma^2 < \sigma_0^2$ use una prueba de una cola situando toda el área α en la cola apropiada de la distribución (superior o inferior repectivamente).

Se ilustra la prueba de hipótesis para la varianza con un ejemplo.

Ejemplo 9.4

Un fabricante de cemento afirma que su cemento posee una resistencia relativamente estable, y que la resistencia, medida en kilogramos por centímetro cuadrado, tiene un recorrido de 40 kilogramos por centímetro cuadrado alrededor del valor medio. De una muestra de $n = 10$ observaciones se obtiene una media y una varianza

$$\bar{y} = 312 \quad \text{y} \quad s^2 = 195$$

¿Presentan estos datos suficiente evidencia para rechazar la afirmación del fabricante?

Solución Como se ha establecido el fabricante afirma que el recorrido de las medidas de resistencia es igual a 40 kilogramos por centímetro cuadrado. Supóngase que lo que quiere decir el fabricante es en realidad que las observaciones caen en este intervalo 95% de las veces y por lo tanto el recorrido es aproximadamente 4σ . Esto es $\sigma = 10$. Entonces se desea probar la hipótesis nula

$$H_0: \sigma^2 = (10)^2 = 100$$

contra la alternativa

$$H_a: \sigma^2 > 100$$

La hipótesis alternativa requiere de una prueba estadística de una sola cola, con la región de rechazo localizada enteramente en la cola superior de la distribución de χ^2 . El valor crítico de χ^2 para $\alpha = .05$ y $n = 10$ es $\chi^2 = 16.9190$, lo que implica que se rechaza H_0 si la estadística de prueba excede este valor.

Haciendo el cálculo se obtiene

$$\chi^2 = \frac{(n - 1)s^2}{\sigma_0^2} = \frac{1755}{100} = 17.55$$

Puesto que el valor de la estadística de prueba cae en la región crítica, se concluye que la hipótesis nula es falsa y que el recorrido de las resistencias del concreto excede al afirmado por el fabricante.

Un intervalo de confianza para σ^2 con confiabilidad de $(1 - \alpha)$ se presenta en el siguiente cuadro.

Intervalo de confianza del $(1 - \alpha)100\%$ para σ^2

$$\frac{(n-1)s^2}{\chi_S^2} < \sigma^2 < \frac{(n-1)s^2}{\chi_I^2}$$

en donde χ_I^2 y χ_S^2 son los valores de χ^2 que localiza un medio de α en la cola inferior y superior respectivamente.

Por ejemplo, un intervalo del 90% de confianza para σ^2 en el ejemplo 9.4 usaría

$$\chi_I^2 = \chi_{.95}^2 = 3.32511$$

$$\chi_S^2 = \chi_{.05}^2 = 16.9190$$

entonces el intervalo de confianza para σ^2 sería

$$\frac{(9)(195)}{16.9190} < \sigma^2 < \frac{(9)(195)}{3.32511}$$

$$103.73 < \sigma^2 < 527.80$$

Ejemplo 9.5 Un investigador está convencido de que su equipo de medición posee una variabilidad medida por una desviación estándar de $\sigma = 2$. Al realizar un experimento él registró observaciones de 4.1, 5.2 y 10.2. ¿Están estos datos en desacuerdo con su suposición? Pruebe la hipótesis $H_0: \sigma = 2$ ó $\sigma^2 = 4$ y establezca un intervalo de confianza del 90% para σ^2 .

Solución La varianza muestral calculada es $s^2 = 10.57$. Puesto que se desea detectar $\sigma^2 > 4$ ó bien $\sigma^2 < 4$, se empleará una prueba de dos colas. Usando $\alpha = .10$ y asignando .05 a cada cola, se rechaza H_0 cuando $\chi^2 > 5.99147$ ó $\chi^2 < .102587$.

El valor calculado para la estadística de prueba es

$$\chi^2 = \frac{(n-1)s^2}{\sigma_0^2} = \frac{(2)(10.57)}{4} = 5.29$$

Puesto que la estadística de prueba no cae en la región de rechazo, los datos no proporcionan evidencia suficiente para rechazar la hipótesis nula $H_0: \sigma^2 = 4$.

El intervalo de confianza correspondiente al 90% es

$$\frac{(n-1)s^2}{\chi_S^2} < \sigma^2 < \frac{(n-1)s^2}{\chi_I^2}$$

con

$$\chi_I^2 = \chi_{.95}^2 = .102587$$

$$\chi_S^2 = \chi_{.05}^2 = 5.99147$$

Substituyendo en la fórmula se tiene

$$\frac{2(10.57)}{5.99147} < \sigma^2 < \frac{2(10.57)}{.102587}$$

$$3.53 < \sigma^2 < 206.1$$

Entonces se estima que la varianza de la población cae en el intervalo de 3.53 a 206.1, el cual resulta un intervalo bastante grande que refleja qué tan pobre es la información acerca de la varianza de la población cuando se cuenta sólo con 3 observaciones. En consecuencia no resulta sorprendente que la evidencia resulte insuficiente para rechazar la hipótesis nula $\sigma^2 = 4$. Para obtener más información sobre σ^2 , el experimentador necesita aumentar el tamaño de la muestra.

Ejercicios

9.24. Muchos fabricantes de muebles para cocina han cambiado sus procedimientos de remachado por puntos de soldadura, en un afán de reducir los costos de manufactura. Sin embargo, resulta importante seguir manteniendo la resistencia al esfuerzo aún con el uso de soldadura. Un fabricante de soldadoras afirma que uno de sus productos puede producir puntos de soldadura en los muebles de cocina y que estos tengan una resistencia al esfuerzo de 400 a 500 kilogramos. Una muestra de $n = 25$ puntos de soldadura producidos por la soldadora en cuestión fue sometida a una prueba de resistencia. La media y la desviación estándar de los esfuerzos soportados en la prueba fueron de $\bar{y} = 438$ kilos y $s = 29$ kilos. ¿Presentan estos datos evidencia suficiente para rechazar la afirmación del fabricante de soldadoras? Use $\alpha = .05$.

9.25. Refiérase al ejercicio 9.24.

a. Estime al esfuerzo medio que pueden soportar los puntos de soldadura producidos por la soldadora del fabricante, usando un intervalo de confianza del 90%.

b. Encuentre un intervalo de confianza del 90% para la varianza de los esfuerzos soportados por los puntos de soldadura.

9.26. Los instrumentos de precisión, como los diseñados para medir volumen, temperatura y contenido

de grasa en la carne, están diseñados para proporcionar una medición que es no sólo correcta en promedio sino que posee también muy poca variación alrededor del valor verdadero. Como un ejemplo elemental considere un medidor de volumen en la bomba de una estación de gasolina. El operador de la estación insiste en que el medidor no subestime el volumen despachado, en tanto el comprador sugiere que no se registre en exceso el volumen despachado. Suponga que un fabricante afirma que su medidor de volumen registra con una precisión de 0.1 litro la cantidad despachada. El medidor del fabricante se coloca en una bomba de gasolina y se despachan $n = 5$ mediciones de exactamente 10 litros usando el medidor. Los volúmenes registrados en el medidor para los cinco ensayos son 10.05, 10.00, 9.90, 9.95, 10.15. ¿Considera usted que estos datos confirman o refutan la afirmación del fabricante? Realice una prueba al $\alpha = .05$ de nivel de significancia. (*Sugerencia:* Si la afirmación del fabricante es de que su producto es preciso hasta 0.1 litro de la verdadera cantidad, está afirmando que la cantidad registrada no distará más de .1 en cualquier dirección de la cantidad despachada. Es decir lo que afirma él es que el rango de lecturas para una cantidad despachada no será mayor a 0.2 litros.)

9.7 Comparación de las varianzas de dos poblaciones

La necesidad de métodos estadísticos para comparar la varianza de dos poblaciones resulta evidente de la discusión en la sección 9.6. Frecuentemente se

desea comparar la precisión de un instrumento de medición con la de otro, la estabilidad de un proceso de manufactura con la de otro, o la variabilidad del sistema de evaluación de un profesor con el de otro.

Intuitivamente, parece razonable comparar las varianzas de dos poblaciones mediante el uso del cociente de las varianzas muestrales s_1^2/s_2^2 . Si este cociente es cercano a 1, se tendrá poca evidencia que indicara que σ_1^2 y σ_2^2 son distintos. Por otra parte un valor o muy grande o muy pequeño de s_1^2/s_2^2 será evidencia de que existe una diferencia en las varianzas de la población.

Pero, ¿qué tan grande o qué tan pequeño debe ser s_1^2/s_2^2 para proporcionar evidencia suficiente para rechazar la hipótesis nula $H_0: \sigma_1^2 = \sigma_2^2$? La respuesta a esta pregunta puede obtenerse al estudiar la distribución muestral de s_1^2/s_2^2 .

Cuando se tienen muestras aleatorias independientes de dos poblaciones normales con varianzas iguales, es decir $\sigma_1^2 = \sigma_2^2$, entonces s_1^2/s_2^2 posee una distribución muestral que es conocida por los estadísticos como la **distribución F**. No es necesario preocuparse por la ecuación para la distribución F . Tal vez baste afirmar que, como podría imaginarse, resulta razonablemente compleja. Será suficiente saber que la distribución es bien conocida y que se cuenta con los valores críticos tabulados. Estos se presentan en las tablas 6 y 7 del apéndice.

La forma de la distribución F no es simétrica y depende de los grados de libertad asociados a s_1^2 y s_2^2 . Estas cantidades se representan por ν_1 y ν_2 respectivamente. Este hecho complica la tabulación de los valores críticos de la distribución F y se requiere de la construcción de una tabla para cada valor que se desee para una área bajo la curva. Las tablas 6 y 7 del apéndice presentan los valores críticos correspondientes a $\alpha = .05$ y $\alpha = .01$, respectivamente. En la figura 9.4 se muestra una distribución F con $\nu_1 = 10$ y $\nu_2 = 10$.

Por ejemplo, la tabla 6 del apéndice presenta los valores de F tales que la probabilidad de que F exceda a $F_{.05}$ es de .05. O dicho de otra manera el 5% del área bajo la distribución F se encuentra a la derecha de $F_{.05}$ (ver figura 9.4). Los grados de libertad ν_1 para s_1^2 se indican en la fila superior de la tabla y los grados de libertad ν_2 para s_2^2 aparecen en la primera columna de la izquierda (ver tabla 9.6).

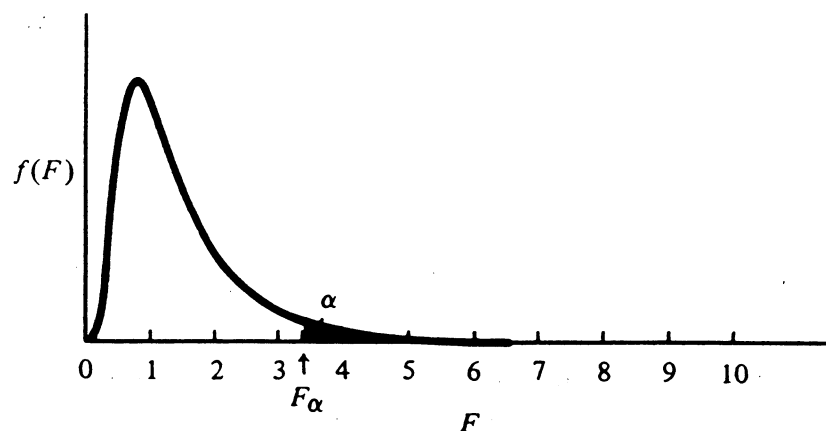


Figura 9.4 Distribución F con $\nu_1 = 10$ y $\nu_2 = 10$

Tabla 9.6 Formato de la tabla F, tabla 6 del apéndice, $\alpha = .05$

GRADOS DE LIBERTAD DEL NUMERADOR															
$\nu_1 \backslash \nu_2$	1	2	3	4	5	6	7	8	9	...	60	120	∞	ν_1	
Grados de libertad del denominador	1	161.4	199.5	215.7	224.6	230.2	234.0	236.8	238.9	240.5	...	252.2	253.3	254.3	1
	2	18.51	19.00	19.16	19.25	19.30	19.33	19.35	19.37	19.38	...	19.48	19.49	19.50	2
	3	10.13	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81	...	8.57	8.55	8.53	3
	4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00	...	5.69	5.66	5.63	4
	5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77	...	4.43	4.40	4.36	5
	6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10	...	3.74	3.70	3.67	6
	7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68	...	3.30	3.27	3.23	7
	8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39	...	3.01	2.97	2.93	8
	9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18	...	2.79	2.75	2.71	9
	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:
	15	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59	...	2.16	2.11	2.07	15
	16	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54	...	2.11	2.06	2.01	16
	17	4.45	3.59	3.20	2.96	2.81	2.70	2.61	2.55	2.49	...	2.06	2.01	1.96	17
	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:

Con referencia a la tabla 9.6, note que la $F_{.05}$ para tamaños de muestra $n_1 = 7$ y $n_2 = 10$ (es decir $\nu_1 = 6$ y $\nu_2 = 9$) es 3.37; asimismo, el valor crítico $F_{.05}$ para tamaños de muestra $n_1 = 9$ y $n_2 = 16$ ($\nu_1 = 8$, $\nu_2 = 15$) es 2.64. Estas cifras han sido marcadas en la tabla 9.6.

De manera similar los valores críticos para $\alpha = .01$ se presentan en la tabla 7 del apéndice. Entonces

$$P(F > F_{.01}) = .01$$

La prueba estadística de la hipótesis nula

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

utiliza la estadística de prueba

$$F = \frac{s_1^2}{s_2^2}$$

Cuando la hipótesis alternativa implica una prueba de una cola, ésto es

$$H_a: \sigma_1^2 > \sigma_2^2$$

las tablas pueden usarse directamente. Sin embargo, cuando la hipótesis alternativa requiere de una prueba de 2 colas, ésto es

$$H_a: \sigma_1^2 \neq \sigma_2^2$$

entonces la región de rechazo será dividida en las colas superior e inferior de la distribución de F y las tablas para los valores críticos de la cola inferior han sido deliberadamente omitidos. La razón de tal omisión tiene una explicación sencilla.

Se tiene siempre la libertad de identificar a cualquiera de las dos poblaciones como la población 1. Si la población con la varianza muestral mayor se denota como población 2, entonces $s_2^2 > s_1^2$ y se requerirá de un rechazo en la cola inferior de la distribución F . Puesto que la identificación de la población es arbitraria se puede evitar esta dificultad si se identifica como población 1 a

la población con mayor varianza muestral. En otras palabras, la varianza muestral mayor irá siempre en el numerador de

$$F = \frac{s_1^2}{s_2^2}$$

e identificará a la población 1. Entonces puesto que el área en la cola derecha representa sólo el $\alpha/2$, este valor debe duplicarse para obtener el valor correcto para la probabilidad α de un error del tipo I. Por lo tanto, si se usa la tabla 6 para una prueba de dos colas, la probabilidad de error tipo I es $\alpha = .10$.

Prueba de hipótesis acerca de la igualdad de las varianzas de dos poblaciones

Hipótesis nula $H_0: \sigma_1^2 = \sigma_2^2$.

Hipótesis alternativa $H_a: \sigma_1^2 \neq \sigma_2^2$, para una prueba estadística de dos colas.

Estadística de prueba: $F = s_1^2/s_2^2$ con s_1^2 la mayor de las varianzas muestrales.

Región de rechazo: Para una prueba de dos colas, rechace H_0 si $F > F_{\alpha/2}$ donde $F_{\alpha/2}$ se basa en $(n_1 - 1)$ y $(n_2 - 1)$ grados de libertad y el valor $\alpha/2$ se duplica para obtener el valor correcto de la probabilidad α de un error de tipo I. (Si se usa la tabla 6 para una prueba de dos colas, la probabilidad de error tipo I es $\alpha = .10$.)

A continuación se ilustran estas ideas con algunos ejemplos.

Ejemplo 9.6

El riesgo asociado con inversiones alternativas se evalúa generalmente por medio de la varianza de los réditos asociados con cada inversión*. La distribución de los réditos para dos inversiones alternativas A y B se muestran en la figura 9.5. La tasa esperada de redeviabilidad para cada inversión es de 17.8%, pero con base en los réditos observados en los 10 años anteriores para la inversión A y los 8 años anteriores para la inversión B, las varianzas para

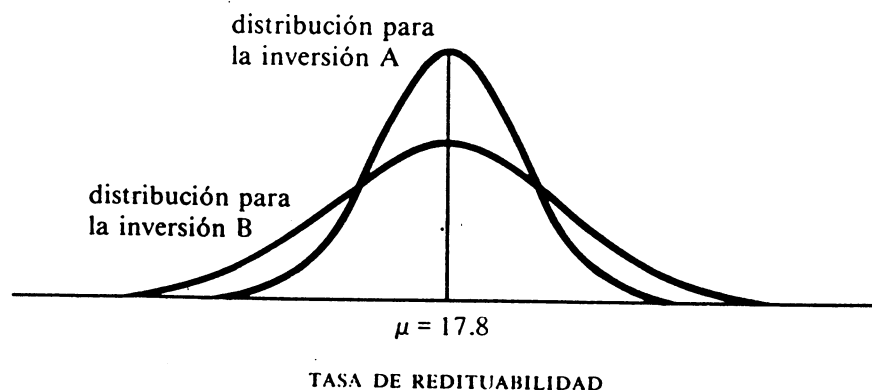


Figura 9.5 Distribución de las tasas de redeviabilidad para las inversiones A y B.

*W.F. Sharpe. *Portfolio Theory and Capital Markets*. (New York: McGraw-Hill, 1970).

los réditos para ambas inversiones son 3.21 y 7.14 respectivamente. ¿Presentan estas varianzas evidencia suficiente para indicar que los riesgos de las inversiones A y B son distintos? (Esto es ¿existe diferencia en las varianzas de las poblaciones?)

Solución Suponga que las poblaciones poseen una distribución de probabilidad de forma razonablemente acampanada y por lo tanto se satisface para propósitos prácticos, la suposición de que las poblaciones son normales.

Se desea probar la hipótesis nula

$$H_0: \sigma_A^2 = \sigma_B^2$$

contra la alternativa

$$H_a: \sigma_A^2 \neq \sigma_B^2$$

Usando la tabla 6, duplicando el área bajo la curva, se rechazará H_0 cuando $F > 3.29$ con $\alpha = .10$, puesto que $n_A = 10$ y $n_B = 8$. (Note que $F_{.05} = 3.29$ se basa en $\nu_1 = 7$ y $\nu_2 = 9$ grados de libertad.)

El valor calculado de la estadística de prueba es

$$F = \frac{s_B^2}{s_A^2} = \frac{7.14}{3.21} = 2.22$$

Nótese que la estadística de prueba no cae en la región de rechazo por lo que no se rechaza $H_0: \sigma_A^2 = \sigma_B^2$, o sea que la evidencia es insuficiente para indicar una diferencia entre las varianzas de las poblaciones.

Ejemplo 9.7

La consistencia en el sabor de la cerveza es una cualidad importante para mantener la lealtad de la clientela. La variabilidad en el sabor de una cerveza dada puede verse afectada por la longitud del período de fermentación, variación en los ingredientes y diferencias en el equipo de fermentación. Un fabricante con dos líneas de producción, 1 y 2, ha hecho ligeros cambios a la línea 2 buscando reducir la variabilidad así como el promedio del índice de sabor. Se toman al azar muestras de $n_1 = 25$ y $n_2 = 25$ vasos de cerveza de cada línea de producción y se determina el índice de sabor con un instrumento apropiado, obteniéndose

$$\bar{y}_1 = 3.2 \quad \bar{y}_2 = 3.0$$

$$s_1^2 = 1.04 \quad s_2^2 = .51$$

¿Presentan estos datos suficiente evidencia para indicar que la variabilidad del proceso es menor para la línea 2? (Esto es, pruebe la hipótesis nula, $H_0: \sigma_1^2 = \sigma_2^2$, en contra la hipótesis alternativa, $H_a: \sigma_1^2 > \sigma_2^2$.)

Solución Para probar la hipótesis nula

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

contra la alternativa

$$H_a: \sigma_1^2 > \sigma_2^2$$

al $\alpha = .05$ de significancia, se rechazará H_0 cuando F sea mayor que $F_{.05} = 1.98$; esto es, se usará una prueba estadística de una sola cola.

El valor calculado de la estadística de prueba,

$$F = \frac{s_1^2}{s_2^2} = \frac{1.04}{.51} = 2.04$$

cae en la región de rechazo, y por lo tanto se concluye que la variabilidad del proceso 2 es menor que la variabilidad del proceso 1.

La prueba de igualdad de las varianzas de dos poblaciones no supone que las medias sean iguales, aunque puede notarse que en ambos ejemplos, 9.6 y 9.7, los promedios son casi iguales. Lo que la prueba examina es la **uniformidad comparativa de los valores individuales dentro de cada población**. Por ejemplo, en el ejemplo 9.6 si se hubiera rechazado la hipótesis nula para concluir que los réditos de una inversión son más variables que los de la otra, se concluiría que aún cuando los réditos promedio de ambas inversiones son casi iguales, una inversión representa mayor riesgo que la otra indicado ésto por su mayor variabilidad. Puesto que la hipótesis de igualdad de varianzas no fue rechazada para ese ejemplo, no existe evidencia suficiente para indicar diferentes riesgos para las dos inversiones. *Nota:* Tal diferencia en riesgos puede existir pero no se obtuvo evidencia suficiente con base en muestras de tamaños $n_A = 10$ y $n_B = 8$.

Ejercicios

9.27. Una planta de producción tiene dos sistemas de fabricación extremadamente complejos, uno de ellos dos veces más viejo que el otro. A ambos sistemas se les lubrica y se les da mantenimiento cada dos semanas. Durante 30 días laborables se registra el número de productos terminados fabricados diariamente por cada uno de los sistemas y se obtienen los siguientes resultados:

SISTEMA NUEVO	SISTEMA VIEJO
$\bar{y}_1 = 246$	$\bar{y}_2 = 240$
$s_1 = 15.6$	$s_2 = 28.2$

¿Presentan estos datos evidencia suficiente para concluir que la variabilidad en la producción diaria justifica un mantenimiento más intensivo para el sistema viejo? Use $\alpha = .05$.

9.28. La estabilidad en las mediciones de las características de un producto manufacturado es importante para mantener la calidad del producto. De hecho, en ocasiones resulta preferible tener poca variación al medir las características importantes del producto y que el promedio esté un poco sesgado a

tener mucha variabilidad y un valor medio correcto. La segunda situación puede producir mayor porcentaje de productos defectuosos que la primera. Un fabricante de bombillas eléctricas sospecha que una de sus líneas de producción está produciendo bombillas con mayor variación en la longitud de vida. Para probar su suposición, compara $n = 50$ bombillas seleccionadas aleatoriamente de la línea sospechosa y $n = 50$ de la línea que parece bajo control. En la tabla se dan las medias y las varianzas de ambas muestras.

LÍNEA SOSPECHOSA	LÍNEA BAJO CONTROL
$\bar{y}_1 = 1,520$	$\bar{y}_2 = 1,476$
$s_1^2 = 92,000$	$s_2^2 = 37,000$

¿Considera que estos datos presentan suficiente evidencia para indicar que las bombillas producidas por la línea sospechosa poseen una mayor varianza en longitud de vida que aquellos producidos por la línea que se supone bajo control? Use $\alpha = .05$.

9.29. En la industria de manufacturas metálicas, la productividad, y consecuentemente la utilidad, dependen grandemente de la calidad y uniformidad de las materias primas necesarias. Suponga que se tienen bajo consideración dos fuentes principales de materia prima, ambas fuentes parecen tener características de control similares, pero el fabricante no está seguro acerca de su respectiva uniformidad en el contenido de impurezas. Se toman 10 muestras de 100 kg cada una de cada fuente y se determina la cantidad de impurezas en cada muestra. Los resultados aparecen en la tabla. ¿Sugieren estos datos la existencia de una diferencia en la uniformidad de contenido de impu-

rezas en los dos materiales? Pruebe con un 10% de significancia.

MATERIAL A	MATERIAL B
$\bar{y}_1 = 41.3$	$\bar{y}_2 = 39.6$
$s_1^2 = 18.75$	$s_2^2 = 7.85$

9.30. Continuando con el ejercicio 9.29, pruebe si existe diferencia en el contenido medio de impurezas de las muestras de los dos materiales. Use $\alpha = .05$.

9.8 Resumen

Es importante notar que las estadísticas t , χ^2 y F usadas en los métodos estadísticos con muestras pequeñas discutidos en las secciones precedentes, se basan en la suposición de que *las poblaciones muestreadas poseen una distribución de probabilidad normal*. Este requerimiento es satisfecho para un buen número de observaciones experimentales.

Es notorio el parecido entre la t de Student y la estadística z y por lo tanto la similitud de los métodos para probar hipótesis y construir intervalos de confianza. Las estadísticas χ^2 y F usadas para hacer inferencias acerca de las varianzas de las poblaciones no siguen este mismo patrón, pero el razonamiento usado en la construcción de pruebas estadísticas y de intervalos de confianza es el mismo para todos los métodos que se han presentado.

Ejercicios complementarios

9.31. ¿Qué suposiciones se hacen cuando se usa una prueba t de Student para probar una hipótesis acerca de la media de una población? ¿Cuáles serán los resultados si estas suposiciones no son válidas pero de todos modos se usa una prueba t de Student?

9.32. Un cierto proceso químico ha producido un promedio de 800 tons de producto por día. Los rendimientos diarios de la semana anterior son 785, 805, 790, 793 y 802 tons. ¿Indican estos datos que el rendimiento promedio es menor que 800 tons y entonces algo malo está pasando con el proceso? Pruebe al 5% de significancia.

9.33. Encuentre un intervalo del 90% de confianza para el rendimiento medio del ejercicio 9.32.

9.34. Refiérase a los ejercicios 9.32 y 9.33. ¿Qué tan grande debe ser la muestra para lograr que la longitud del intervalo de confianza se reduzca a 5 tons?

9.35. Se tomó una muestra aleatoria de $n = 20$ secretarías empleadas en agencias públicas de cierta ciudad y se encontró en esta muestra que el salario promedio por hora era de $\bar{y} = \$3.95$, con desviación estándar de \$0.58. Encuentre un intervalo de confianza de 95% para el salario medio por hora para las secretarías empleadas en las agencias públicas de la ciudad.

9.36. Una máquina para vender bebidas automáticamente está diseñada para despachar, en promedio, 7 onzas de bebida por cada vaso. Para probar la máquina se tomaron de ésta 10 vasos de bebida y se midió su contenido. La media y la desviación estándar de los 10 vasos fue de $\bar{y} = 7.1$ onzas y $s^2 = 0.12$ onzas respectivamente. ¿Presentan estos datos evidencia para indicar que la cantidad media despachada difiere de 7 onzas? Pruebe con un nivel de significancia del 10%.