Intervalo de predicción de una observación futura cuando se desconoce σ^2

Para una distribución normal de mediciones cuando la media μ y la varianza σ^2 se desconocen, un **intervalo de predicción** del 100(1 – α)% de una observación futura x_0 es

$$\bar{x} - t_{\alpha/2}s\sqrt{1 + 1/n} < x_0 < \bar{x} + t_{\alpha/2}s\sqrt{1 + 1/n},$$

donde $t_{\alpha/2}$ es el valor t con $v = n - 1$ grados de libertad, que deja una área de $\alpha/2$ a la derecha.

También se pueden utilizar intervalos de predicción unilaterales. Los límites de predicción superiores se aplican en casos en los que es necesario enfocarse en observaciones futuras grandes. El interés por observaciones pequeñas futuras requiere utilizar límites de predicción más bajos. El límite superior es dado por

$$\bar{x} + t_{\alpha}s\sqrt{1 + 1/n}$$

y el límite inferior por

$$\bar{x} - t_{\alpha}s\sqrt{1 + 1/n}.$$

Ejemplo 9.8: Un inspector de alimentos seleccionó aleatoriamente 30 paquetes de carne de res 95% magra. La muestra dio como resultado una media de 96.2% con una desviación estándar muestral de 0.8%. Calcule un intervalo de predicción del 99% para la condición baja en grasa de un paquete nuevo. Suponga normalidad.

***Solución*:** Para $v = 29$ grados de libertad, $t_{0.005} = 2.756$. Por lo tanto, un intervalo de predicción del 99% para una observación nueva x_0 es

$$96.2 - (2.756)(0.8)\sqrt{1 + \frac{1}{30}} < x_0 < 96.2 + (2.756)(0.8)\sqrt{1 + \frac{1}{30}},$$

que se reduce a (93.96, 98.44).

Uso de límites de predicción para detectar valores extremos

Hasta el momento hemos puesto poca atención al concepto de **valores extremos** u observaciones aberrantes. La mayoría de los investigadores científicos son muy sensibles a la existencia de observaciones de valores extremos, también llamados datos defectuosos o “malos”. En el capítulo 12 profundizaremos en el estudio de este concepto. Sin embargo, nos interesa considerarlos aquí porque la detección de los valores extremos está estrechamente relacionada con los intervalos de predicción.

Para nuestros propósitos nos conviene considerar que una observación extrema es una que proviene de una población con una media diferente a la que determina el resto de la muestra de tamaño n que se está estudiando. El intervalo de predicción produce un límite que “cubre” una sola observación futura con probabilidad 1 – α, si ésta proviene de la población de la que se tomó la muestra. Por lo tanto, una metodología para detectar valores extremos implica la regla de que **una observación es un valor extremo si cae fuera del intervalo de predicción calculado sin incluir la observación cuestionable en la muestra**. Como resultado, para el intervalo de predicción del ejemplo 9.8, en el caso de los paquetes de carne, la observación que se obtiene al medir un nuevo paquete y encontrar que su contenido libre de grasa está fuera del intervalo (93.96, 98.44) se podría considerar como un valor extremo.

9.7 Límites de tolerancia

Como vimos en la sección 9.6, el científico o el ingeniero podrían estar menos interesados en estimar parámetros que en obtener información sobre el lugar en el que caería una *observación* o medición individual. Este tipo de situaciones requiere intervalos de predicción. Sin embargo, existe un tercer tipo de intervalo que es útil en muchas aplicaciones. Una vez más, suponga que el interés se centra en torno a la fabricación de la pieza de un componente y que existen especificaciones sobre una dimensión de esa parte. Además, la media de esa dimensión no es tan importante. Sin embargo, a diferencia del escenario de la sección 9.6, se podría estar menos interesado en una sola observación y más en el lugar en el que cae la mayoría de la población. Si las especificaciones del proceso son importantes, el administrador del proceso se interesará en el desempeño a largo plazo, **no en la siguiente observación**. Debemos tratar de determinar los límites que, en cierto sentido probabilístico, "cubren" los valores en la población, es decir, los valores medidos de la dimensión.

Un método para establecer el límite deseado consiste en determinar un intervalo de confianza sobre una *proporción fija* de las mediciones. Esto se comprende mejor visualizando una situación en la que se realiza un muestreo aleatorio de una distribución normal con media conocida μ y varianza σ^2. Evidentemente, un límite que cubre el 95% central de la población de observaciones es

$$\mu \pm 1.96\sigma.$$

A esto se le llama **intervalo de tolerancia** y, en realidad, su cobertura del 95% de las observaciones medidas es exacta. Sin embargo, en la práctica rara vez se conocen μ y σ; por consiguiente, el usuario debe aplicar

$$\bar{x} \pm ks.$$

Ahora bien, el intervalo es, desde luego, una variable aleatoria, por lo tanto, la *cobertura* de una proporción de la población por el intervalo no es exacta. Como resultado, se debe usar un intervalo de confianza del $100(1-\gamma)\%$, ya que no se puede esperar que $\bar{x} \pm ks$ cubra cualquier proporción específica todo el tiempo. Lo anterior nos lleva a la siguiente definición.

Límites de tolerancia

Para una distribución normal de mediciones en la que se desconoce la media μ y la desviación estándar σ, los **límites de tolerancia** son dados por $\bar{x} \pm ks$, donde k se determina de tal manera que se pueda estar seguro, con un $100(1-\gamma)\%$ de confianza, de que los límites dados contienen al menos la proporción $1-\alpha$ de las mediciones.

La tabla A.7 ofrece valores de k para $1-\alpha = 0.90, 0.95, 0.99$; $\gamma = 0.05, 0.01$; y para valores seleccionados de n de 2 a 300.

Ejemplo 9.9: Considere el ejemplo 9.8. Con la información dada calcule un intervalo de tolerancia que proporcione límites bilaterales del 95% sobre el 90% de la distribución de paquetes de carne 95% magra. Suponga que los datos provienen de una distribución aproximadamente normal.

***Solución*:** Del ejemplo 9.8, recuerde que $n = 30$, que la media muestral es de 96.2% y que la desviación estándar muestral es de 0.8%. De la tabla A.7, $k = 2.14$. Si utilizamos

$$\bar{x} \pm ks = 96.2 \pm (2.14)(0.8),$$

encontramos que los límites inferior y superior son de 94.5 y de 97.9.

Tenemos 95% de confianza en que el rango anterior cubre el 90% central de la distribución de paquetes de carne de res 95% magra.

Diferencia entre intervalos de confianza, intervalos de predicción e intervalos de tolerancia

Es importante resaltar la diferencia entre los tres tipos de intervalos que se estudiaron e ilustraron en las secciones anteriores. Los cálculos son sencillos, pero la interpretación podría resultar confusa. En aplicaciones de la vida real tales intervalos no son intercambiables, ya que sus interpretaciones son muy diferentes.

En el caso de los intervalos de confianza sólo se pone atención en la **media de la población**. Por ejemplo, el ejercicio 9.13 de la página 283 se refiere a un proceso de ingeniería que produce alfileres para costura. Se establece una especificación sobre la dureza de Rockwell por debajo de la cual el cliente no aceptará ningún alfiler. En este caso un parámetro de la población debe tener poca relevancia. Es importante que el ingeniero sepa en dónde *van a estar la mayoría de los valores de la dureza de Rockwell*. Por consiguiente, se deberían utilizar los límites de tolerancia. Seguramente, al administrador le agradará saber que los límites de tolerancia en cualquier producto del proceso son más rigurosos que las especificaciones para el propio proceso.

Es verdad que la interpretación del límite de tolerancia se relaciona hasta cierto punto con el intervalo de confianza. El intervalo de tolerancia del $100(1 - \alpha)\%$ sobre, digamos, la proporción 0.95, se podría considerar como un intervalo de confianza **sobre el 95% intermedio** de la distribución normal correspondiente. Los límites de tolerancia unilaterales también son relevantes. En el caso del problema de dureza de Rockwell se desearía tener un límite inferior de la forma $\bar{x} - ks$, tal que se tenga un 99% de confianza en que al menos 99% de los valores de la dureza de Rockwell excederán al valor calculado.

Los intervalos de predicción se pueden aplicar cuando es importante determinar un límite para un **solo valor**. Aquí la media no es la cuestión, ni tampoco la ubicación de la mayoría de la población, lo que se requiere, más bien, es la ubicación de una sola nueva observación.

Estudio de caso 9.1: **Calidad de una máquina.** Una máquina produce piezas de metal que tienen forma cilíndrica. Se toma una muestra de tales piezas y se encuentra que los diámetros son 1.01, 0.97. 1.03, 1.04, 0.99, 0.98, 0.99, 1.01 y 1.03 centímetros. Utilice estos datos para calcular tres tipos de intervalos y hacer interpretaciones que ilustren las diferencias entre ellos en el contexto del sistema. Para todos los cálculos suponga una distribución aproximadamente normal. La media muestral y la desviación estándar para los datos dados son $\bar{x} = 1.0056$ y $s = 0.0246$.

a) Calcule un intervalo de confianza del 99% sobre la media del diámetro.

b) Calcule un intervalo de predicción del 99% sobre el diámetro medido de una sola pieza de metal tomada de la máquina.

c) Calcule los límites de tolerancia del 99% que contengan 95% de las piezas de metal producidas por esta máquina.

***Solución*:** *a)* El intervalo de confianza del 99% para la media del diámetro está dado por

$$\bar{x} \pm t_{0.005}\, s/\sqrt{n} = 1.0056 \pm (3.355)(0.0246/3) = 1.0056 \pm 0.0275.$$

Por lo tanto, los límites de confianza del 99% son 0.9781 y 1.0331.

b) El intervalo de predicción del 99% para una futura observación está dado por

$$\bar{x} \pm t_{0.005}\, s\sqrt{1 + 1/n} = 1.0056 \pm (3.355)(0.0246)\sqrt{1 + 1/9},$$

donde los límites son 0.9186 y 1.0926.

c) De la tabla A.7, para $n = 9$, $1 - \gamma = 0.99$, y $1 - \alpha = 0.95$, obtenemos $k = 4.550$ para los límites bilaterales. Por lo tanto, los límites de tolerancia del 99% son dados por

$$\bar{x} + ks = 1.0056 \pm (4.550)(0.0246),$$

donde los límites son 0.8937 y 1.1175. Tenemos un 99% de confianza en que el intervalo de tolerancia de 0.8937 a 1.1175 contendrá el 95% central de la distribución de diámetros producidos.

Este estudio de caso ilustra que los tres tipos de límites pueden conducir a resultados muy diferentes, aunque todos son límites del 99%. En el caso del intervalo de confianza sobre la media, el 99% de estos intervalos cubre la media del diámetro de la población. Por lo tanto, decimos que tenemos un 99% de confianza en que la media del diámetro producido por el proceso se encuentra entre 0.9781 y 1.0331 centímetros. Se hace hincapié en la media y se pone poco interés en una sola lectura o en la naturaleza general de la distribución de diámetros en la población. En lo que se refiere a los límites de predicción, los límites 0.9186 y 1.0926 se basan en la distribución de una sola pieza "nueva" de metal tomada del proceso, y nuevamente el 99% de estos límites cubren el diámetro de una nueva pieza medida. Por otro lado, como se sugirió en la sección anterior, los límites de tolerancia le dan al ingeniero una idea de en qué parte de la población se localiza la "mayoría", digamos el 95% central, de los diámetros de las piezas medidas. Los límites de tolerancia del 99%, 0.8937 y 1.1175 difieren mucho de los otros dos límites. Si esos límites le parecen demasiado anchos al ingeniero, esto se reflejará de forma negativa en la calidad del proceso. Por otro lado, si los límites representan un resultado deseable, el ingeniero podría concluir que la mayoría (95% en este caso) de los diámetros se encuentran dentro de un rango adecuado. De nuevo, se podría hacer una interpretación del intervalo de confianza, a saber, el 99% de esos límites calculados cubrirán el 95% intermedio de la población de diámetros.

Ejercicios

9.1 Un investigador de la UCLA afirma que la esperanza de vida de los ratones se puede extender hasta en 25% cuando se reduce aproximadamente 40% de las calorías de su dieta desde el momento en que son destetados. La dieta restringida se enriquece hasta niveles normales con vitaminas y proteínas. Si se supone que a partir de estudios previos se sabe que $\sigma = 5.8$ meses, ¿cuántos ratones se deberían incluir en la muestra para tener un 99% de confianza en que la vida media esperada de la muestra estará dentro de 2 meses a partir de la media de la población para todos los ratones sujetos a la dieta reducida?

9.2 Una empresa de material eléctrico fabrica bombillas que tienen una duración distribuida de forma aproximadamente normal, con una desviación estándar de 40 horas. Si una muestra de 30 bombillas tiene una duración promedio de 780 horas, calcule un intervalo de confianza del 96% para la media de la población de todas las bombillas producidas por esta empresa.

9.3 Muchos pacientes con problemas del corazón tienen un marcapasos para controlar su ritmo cardiaco. El marcapasos tiene montado un módulo conector de plástico en la parte superior. Suponga una desviación estándar de 0.0015 pulgadas y una distribución aproximadamente normal, y con base en esto calcule un intervalo de confianza del 95% para la media de la profundidad de todos los módulos conectores fabricados por cierta empresa. Una muestra aleatoria de 75 módulos tiene una profundidad promedio de 0.310 pulgadas.

9.4 Las estaturas de una muestra aleatoria de 50 estudiantes universitarios tienen una media de 174.5 centímetros y una desviación estándar de 6.9 centímetros.

a) Construya un intervalo de confianza del 98% para la estatura media de todos los estudiantes universitarios.

b) ¿Qué podemos afirmar con una confianza del 98% acerca del posible tamaño de nuestro error, si estimamos que la estatura media de todos los estudiantes universitarios es de 174.5 centímetros?

9.5 Una muestra aleatoria de 100 propietarios de automóviles del estado de Virginia revela que éstos conducen su automóvil, en promedio, 23,500 kilómetros por año, con una desviación estándar de 3900 kilómetros. Suponga que la distribución de las mediciones es aproximadamente normal.

a) Construya un intervalo de confianza del 99% para el número promedio de kilómetros que un propietario de un automóvil conduce anualmente en Virginia.

b) ¿Qué podemos afirmar con un 99% de confianza acerca del posible tamaño del error, si estimamos que los propietarios de automóviles de Virginia conducen un promedio de 23,500 kilómetros por año?

9.6 ¿Qué tan grande debe ser la muestra en el ejercicio 9.2 si deseamos tener un 96% de confianza en que nuestra media muestral estará dentro de 10 horas a partir de la media verdadera?

9.7 ¿De qué tamaño debe ser la muestra en el ejercicio 9.3 si deseamos tener un 95% de confianza en que nuestra media muestral estará dentro de un 0.0005 de pulgada de la media verdadera?

9.8 Un experto en eficiencia desea determinar el tiempo promedio que toma perforar tres hoyos en cierta placa metálica. ¿De qué tamaño debe ser una muestra para tener un 95% de confianza en que esta media muestral estará dentro de 15 segundos de la media verdadera? Suponga que por estudios previos se sabe que $\sigma = 40$ segundos.

9.9 Según estudios realizados por el doctor W. H. Bowen, del Instituto Nacional de Salud, y por el doctor J. Yudben, profesor de nutrición y dietética de la Universidad de Londres, el consumo regular de cereales preendulzados contribuye al deterioro de los dientes, a las enfermedades cardiacas y a otras enfermedades degenerativas. En una muestra aleatoria de 20 porciones sencillas similares del cereal Alpha-Bits, el contenido promedio de azúcar era de 11.3 gramos con una desviación estándar de 2.45 gramos. Suponga que el contenido de azúcar está distribuido normalmente y con base en esto construya un intervalo de confianza de 95% para el contenido medio de azúcar de porciones sencillas de Alpha-Bits.

9.10 Las integrantes de una muestra aleatoria de 12 graduadas de cierta escuela para secretarias teclearon un promedio de 79.3 palabras por minuto, con una desviación estándar de 7.8 palabras por minuto. Suponga una distribución normal para el número de palabras que teclean por minuto y con base en esto calcule un intervalo de confianza del 95% para el número promedio de palabras que teclean todas las graduadas de esta escuela.

9.11 Una máquina produce piezas metálicas de forma cilíndrica. Se toma una muestra de las piezas y los diámetros son 1.01, 0.97, 1.03, 1.04, 0.99, 0.98, 0.99, 1.01 y 1.03 centímetros. Calcule un intervalo de confianza del 99% para la media del diámetro de las piezas que se manufacturan con esta máquina. Suponga una distribución aproximadamente normal.

9.12 Una muestra aleatoria de 10 barras energéticas de chocolate de cierta marca tiene, en promedio, 230 calorías por barra y una desviación estándar de 15 calorías. Construya un intervalo de confianza del 99% para el contenido medio verdadero de calorías de esta marca de barras energéticas de chocolate. Suponga que la distribución del contenido calórico es aproximadamente normal.

9.13 En un estudio para determinar la dureza de Rockwell en la cabeza de alfileres para costura se toma una muestra aleatoria de 12. Se toman mediciones de la dureza de Rockwell para cada una de las 12 cabezas y se obtiene un valor promedio de 48.50, con una desviación estándar muestral de 1.5. Suponga que las mediciones se distribuyen de forma normal y con base en esto construya un intervalo de confianza de 90% para la dureza media de Rockwell.

9.14 Se registran las siguientes mediciones del tiempo de secado, en horas, de cierta marca de pintura vinílica:

3.4	2.5	4.8	2.9	3.6
2.8	3.3	5.6	3.7	2.8
4.4	4.0	5.2	3.0	4.8

Suponga que las mediciones representan una muestra aleatoria de una población normal y con base en esto calcule el intervalo de predicción del 95% para el tiempo de secado de la siguiente prueba de pintura.

9.15 Remítase al ejercicio 9.5 y construya un intervalo de predicción del 99% para los kilómetros que viaja anualmente el propietario de un automóvil en Virginia.

9.16 Considere el ejercicio 9.10 y calcule el intervalo de predicción del 95% para el siguiente número observado de palabras por minuto tecleadas por una graduada de la escuela de secretarias.

9.17 Considere el ejercicio 9.9 y calcule un intervalo de predicción del 95% para el contenido de azúcar de la siguiente porción de cereal Alpha-Bits.

9.18 Remítase al ejercicio 9.13 y construya un intervalo de tolerancia del 95% que contenga el 90% de las mediciones.

9.19 Una muestra aleatoria de 25 tabletas de aspirina con antiácido contiene, en promedio, 325.05 mg de aspirina en cada tableta, con una desviación estándar de 0.5 mg. Calcule los límites de tolerancia del 95% que contendrán 90% del contenido de aspirina para esta marca. Suponga que el contenido de aspirina se distribuye normalmente.

9.20 Considere la situación del ejercicio 9.11. Aunque la estimación de la media del diámetro es importante, no es ni con mucho tan importante como intentar determinar la ubicación de la mayoría de la distribución de los diámetros. Calcule los límites de tolerancia del 95% que contengan el 95% de los diámetros.

9.21 En un estudio realizado por el Departamento de Zoología del Virginia Tech con el fin de conocer la cantidad de ortofósforo en el río, se recolectaron 15 "muestras" de agua en una determinada estación ubicada en el río James. La concentración del químico se midió en miligramos por litro. Suponga que la media en la estación de muestreo no es tan importante como la distribución de las concentraciones del químico en los extremos superiores. El interés se centra en saber si las concentraciones en estos extremos son demasiado elevadas. Las lecturas de las 15 muestras de agua proporcionaron una media muestral de 3.84 miligramos por litro y una desviación estándar muestral de 3.07 miligramos por litro. Suponga que las lecturas son una muestra aleatoria de una distribución normal. Calcule un intervalo de predicción (límite de predicción superior del 95%) y un límite de tolerancia (un límite de tolerancia superior del 95% que excede al 95% de la población de valores). Interprete ambos límites, es decir, especifique qué indica cada uno acerca de los extremos superiores de la distribución de ortofósforo en la estación de muestreo.

9.22 Se están estudiando las propiedades de resistencia a la tensión de un determinado tipo de hilo. Con ese fin se prueban 50 piezas en condiciones similares y los resultados que se obtienen revelan una resistencia a la tensión promedio de 78.3 kilogramos y una desviación estándar de 5.6 kilogramos. Suponga que la resistencia a la tensión tiene una distribución normal y con base en esto calcule un límite de predicción inferior al 95% de un solo valor observado de resistencia a la tensión. Además, determine un límite inferior de tolerancia del 95% que sea excedido por el 99% de los valores de resistencia a la tensión.

9.23 Remítase al ejercicio 9.22. ¿Por qué las 1/2 cantidades solicitadas en el ejercicio parecen ser más importantes para el fabricante del hilo que, por ejemplo, un intervalo de confianza en la resistencia media a la tensión?

9.24 Remítase una vez más al ejercicio 9.22. Suponga que un comprador del hilo especifica que éste debe tener una resistencia a la tensión de por lo menos 62 kilogramos. El fabricante estará satisfecho si la cantidad de piezas producidas que no cumplen la especificación no excede al 5%. ¿Hay alguna razón para preocuparse? Esta vez utilice un límite de tolerancia unilateral del 99% que sea excedido por el 95% de los valores de resistencia a la tensión.

9.25 Considere las mediciones del tiempo de secado del ejercicio 9.14. Suponga que las 15 observaciones en el conjunto de datos también incluyen un decimosexto valor de 6.9 horas. En el contexto de las 15 observaciones originales, ¿el valor decimosexto es un valor extremo? Muestre el procedimiento.

9.26 Considere los datos del ejercicio 9.13. Suponga que el fabricante de los alfileres insiste en que la dureza de Rockwell del producto es menor o igual que 44.0 sólo un 5% de las veces. ¿Cuál es su reacción? Utilice un cálculo de un límite de tolerancia como la base de su veredicto.

9.27 Considere la situación del estudio de caso 9.1 de la página 281, con una muestra más grande de piezas metálicas. Los diámetros son los siguientes: 1.01, 0.97, 1.03, 1.04, 0.99, 0.98, 1.01, 1.03, 0.99, 1.00, 1.00, 0.99, 0.98, 1.01, 1.02, 0.99 centímetros. Nuevamente puede suponer una distribución normal. Haga lo siguiente y compare sus resultados con los del estudio de caso. Analice en qué difieren y por qué.

a) Calcule un intervalo de confianza del 99% de la media del diámetro.

b) Calcule un intervalo de predicción del 99% en la medición del siguiente diámetro.

c) Calcule un intervalo de tolerancia del 99% para la cobertura del 95% central de la distribución de diámetros.

9.28 En la sección 9.3 destacamos el concepto del "estimador más eficaz" comparando la varianza de dos estimadores insesgados $\hat{\Theta}_1$ y $\hat{\Theta}_2$. Sin embargo, esto no toma en cuenta el sesgo en el caso en que uno o ambos estimadores no son sesgados. Considere la cantidad

$$EME = E(\hat{\Theta} - \theta),$$

donde EME denota el **error cuadrático medio**. El error cuadrático medio a menudo se utiliza para comparar dos estimadores $\hat{\Theta}_1$ y $\hat{\Theta}_2$ de θ, cuando uno o ambos no son sesgados porque i) es intuitivamente razonable y ii) se toma en cuenta para el sesgo. Demuestre que el EME se puede escribir como

$$EME = E[\hat{\Theta} - E(\hat{\Theta})]^2 + [E(\hat{\Theta} - \theta)]^2$$
$$= \text{Var}(\hat{\Theta}) + [\text{sesgo}(\hat{\Theta})]^2.$$

9.29 Definamos $S'^2 = \sum_{i=1}^{n} (X_i - \bar{X})^2/n$. Demuestre que

$$E(S'^2) = [(n-1)/n]\sigma^2,$$

y, en consecuencia, que S'^2 es un estimador sesgado para σ^2.

9.30 Considere S'^2, el estimador de σ^2, del ejercicio 9.29. Con frecuencia los analistas utilizan S'^2 en lugar de dividir $\sum_{i=1}^{n}(X_i - \bar{X})^2$ entre $n-1$, los grados de libertad en la muestra.

a) ¿Cuál es el sesgo de S'^2?

b) Demuestre que el sesgo de S'^2 se aproxima a cero a medida que $n \to \infty$.

9.31 Si X es una variable aleatoria binomial, demuestre que

a) $\hat{P} = X/n$ es un estimador insesgado de p;

b) $P' = \frac{X + \sqrt{n}/2}{n + \sqrt{n}}$ es un estimador sesgado de p.

9.32 Demuestre que el estimador P' del ejercicio 9.31*b*) se vuelve no sesgado a medida que $n \to \infty$.

9.33 Compare S^2 y S'^2 (véase el ejercicio 9.29), los dos estimadores de σ^2, para determinar cuál es más eficaz. Suponga que estos estimadores se obtienen usando $X_1, X_2, \ldots, X_n$, las variables aleatorias independientes de $n(x; \mu, \sigma)$. ¿Cuál es el estimador más eficaz si se considera sólo la varianza de los estimadores? [*Sugerencia*: Utilice el teorema 8.4 y el hecho de que la varianza de χ^2_v es $2v$, de la sección 6.7.]

9.34 Considere el ejercicio 9.33. Utilice el *EME* que se estudió en el ejercicio 9.28 para determinar qué estimador es más eficaz. Escriba

$$\frac{EME\ (S^2)}{EME\ (S'^2)}.$$

9.8 Dos muestras: estimación de la diferencia entre dos medias

Si tenemos dos poblaciones con medias μ_1 y μ_2, y varianzas σ_1^2 y σ_2^2, respectivamente, el estadístico que da un estimador puntual de la diferencia entre μ_1 y μ_2 es $\bar{X}_1 - \bar{X}_2$. Por lo tanto, para obtener una estimación puntual de $\mu_1 - \mu_2$, se seleccionan dos muestras aleatorias independientes, una de cada población, de tamaños n_1 y n_2, y se calcula $\bar{x}_1 - \bar{x}_2$, la diferencia de las medias muestrales. Evidentemente, debemos considerar la distribución muestral de $\bar{X}_1 - \bar{X}_2$.

De acuerdo con el teorema 8.3, podemos esperar que la distribución muestral de $\bar{X}_1 - \bar{X}_2$ esté distribuida de forma aproximadamente normal con media $\mu_{\bar{X}_1 - \bar{X}_2} = \mu_1 - \mu_2$ y desviación estándar $\sigma_{\bar{X}_1 - \bar{X}_2} = \sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}$. Por lo tanto, podemos asegurar, con una probabilidad de $1 - \alpha$, que la variable normal estándar

$$Z = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}}$$

caerá entre $-z_{\alpha/2}$ y $z_{\alpha/2}$. Si nos remitimos una vez más a la figura 9.2, escribimos

$$P(-z_{\alpha/2} < Z < z_{\alpha/2}) = 1 - \alpha.$$

Al sustituir para Z, establecemos de manera equivalente que

$$P\left(-z_{\alpha/2} < \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}} < z_{\alpha/2}\right) = 1 - \alpha,$$

que conduce al siguiente intervalo de confianza del $100(1 - \alpha)\%$ para $\mu_1 - \mu_2$.

Intervalo de confianza para $\mu_1 - \mu_2$ cuando se conocen σ_1^2 y σ_2^2

Si $\bar{x}_1$ y $\bar{x}_2$ son las medias de muestras aleatorias independientes de tamaños n_1 y n_2, de poblaciones que tienen varianzas conocidas σ_1^2 y σ_2^2, respectivamente, un intervalo de confianza del $100(1 - \alpha)\%$ para $\mu_1 - \mu_2$ es dado por

$$(\bar{x}_1 - \bar{x}_2) - z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} < \mu_1 - \mu_2 < (\bar{x}_1 - \bar{x}_2) + z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}},$$

donde $z_{\alpha/2}$ es el valor z que deja una área de $\alpha/2$ a la derecha.

El grado de confianza es exacto cuando las muestras se seleccionan de poblaciones normales. Para poblaciones no normales el teorema del límite central permite una buena aproximación para muestras de tamaño razonable.

Las condiciones experimentales y la unidad experimental

Para el caso en que se necesita estimar un intervalo de confianza sobre la diferencia entre dos medias se requiere considerar las condiciones experimentales durante el proceso de recolección de datos. Se supone que tenemos dos muestras aleatorias independientes de distribuciones con medias μ_1 y μ_2, respectivamente. Es importante que las condiciones experimentales se parezcan al ideal descrito por las suposiciones tanto como sea posible. Con mucha frecuencia el experimentador debería planear la estrategia del experimento de acuerdo con esto. Para casi cualquier estudio de este tipo existe una *unidad experimental*, que es la parte del experimento que produce el error experimental y genera la varianza de la población que denominamos σ^2. En un estudio farmacológico la unidad experimental es el paciente o el sujeto. En un experimento de agricultura puede ser una superficie de tierra. En un experimento químico puede ser una cantidad de materias primas. Es importante que las diferencias entre tales unidades tengan un impacto mínimo sobre los resultados. El experimentador tendrá un grado de seguridad de que las unidades experimentales no sesgarán los resultados si las condiciones que definen a las dos poblaciones se *asignan al azar a* las unidades experimentales. En los siguientes capítulos acerca de la prueba de hipótesis nos volveremos a concentrar en la aleatorización.

Ejemplo 9.10: Se llevó a cabo un experimento donde se compararon dos tipos de motores, el A y el B. Se midió el rendimiento de combustible en millas por galón. Se realizaron 50 experimentos con el motor tipo A y 75 con el motor tipo B. La gasolina utilizada y las demás condiciones se mantuvieron constantes. El rendimiento promedio de gasolina para el motor A fue de 36 millas por galón y el promedio para el motor B fue de 42 millas por galón. Calcule un intervalo de confianza del 96% sobre $\mu_B - \mu_A$, donde μ_A y μ_B corresponden a la media de la población del rendimiento de millas por galón para los motores A y B, respectivamente. Suponga que las desviaciones estándar de la población son 6 y 8 para los motores A y B, respectivamente.

***Solución*:** La estimación puntual de $\mu_B - \mu_A$ es $\bar{x}_B - \bar{x}_A = 42 - 36 = 6$. Si usamos $\alpha = 0.04$, obtenemos $z_{0.02} = 2.05$ de la tabla A.3. Por lo tanto, sustituyendo en la fórmula anterior, el intervalo de confianza del 96% es

$$6 - 2.05\sqrt{\frac{64}{75} + \frac{36}{50}} < \mu_B - \mu_A < 6 + 2.05\sqrt{\frac{64}{75} + \frac{36}{50}},$$

o simplemente $3.43 < \mu_B - \mu_A < 8.57$.

Este procedimiento para estimar la diferencia entre dos medias se aplica si se conocen σ_1^2 y σ_2^2. Si las varianzas no se conocen y las dos distribuciones implicadas son aproximadamente normales, la distribución t resulta implicada como en el caso de una sola muestra. Si no se está dispuesto a suponer normalidad, muestras grandes (digamos mayores que 30) permitirán usar s_1 y s_2 en lugar de σ_1 y σ_2, respectivamente, con el fundamento de que $s_1 \approx \sigma_1$ y $s_2 \approx \sigma_2$. De nuevo, por supuesto, el intervalo de confianza es aproximado.

Varianzas desconocidas pero iguales

Considere el caso donde se desconocen σ_1^2 y σ_2^2. Si $\sigma_1^2 = \sigma_2^2 = \sigma^2$ obtenemos una variable normal estándar de la forma

$$Z = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\sigma^2[(1/n_1) + (1/n_2)]}}.$$

De acuerdo con el teorema 8.4, las dos variables aleatorias

$$\frac{(n_1 - 1)S_1^2}{\sigma^2} \quad \text{y} \quad \frac{(n_2 - 1)S_2^2}{\sigma^2}$$

tienen distribuciones chi cuadrada con $n_1 - 1$ y $n_2 - 1$ grados de libertad, respectivamente. Además, son variables chi cuadrada independientes, ya que las muestras aleatorias se seleccionaron de forma independiente. En consecuencia, su suma

$$V = \frac{(n_1 - 1)S_1^2}{\sigma^2} + \frac{(n_2 - 1)S_2^2}{\sigma^2} = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{\sigma^2}$$

tiene una distribución chi cuadrada con $v = n_1 + n_2 - 2$ grados de libertad.

Como se puede demostrar que las expresiones anteriores para Z y V son independientes, del teorema 8.5 se sigue que el estadístico

$$T = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\sigma^2[(1/n_1) + (1/n_2)]}} \bigg/ \sqrt{\frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{\sigma^2(n_1 + n_2 - 2)}}$$

tiene la distribución t con $v = n_1 + n_2 - 2$ grados de libertad.

Se puede obtener una estimación puntual de la varianza común desconocida σ^2 agrupando las varianzas muestrales. Si representamos con S_p^2 al estimador agrupado, obtenemos lo siguiente,

Estimado agrupado de la varianza

$$S_p^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}.$$

Al sustituir S_p^2 en el estadístico T, obtenemos la forma menos engorrosa:

$$T = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{S_p\sqrt{(1/n_1) + (1/n_2)}}.$$

Si usamos el estadístico T, tenemos

$$P(-t_{\alpha/2} < T < t_{\alpha/2}) = 1 - \alpha,$$

donde $t_{\alpha/2}$ es el valor t con $n_1 + n_2 - 2$ grados de libertad, por arriba del cual encontramos una área de $\alpha/2$. Al sustituir por T en la desigualdad, escribimos

$$P\left[-t_{\alpha/2} < \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{S_p\sqrt{(1/n_1) + (1/n_2)}} < t_{\alpha/2}\right] = 1 - \alpha.$$

Después de realizar las manipulaciones matemáticas de costumbre, se calculan la diferencia de las medias muestrales $\bar{x}_1 - \bar{x}_2$ y la varianza agrupada, y se obtiene el siguiente intervalo de confianza del $100(1 - \alpha)\%$ para $\mu_1 - \mu_2$.

Se observa con facilidad que el valor de s_p^2 es un promedio ponderado de las dos varianzas muestrales s_1^2 y s_2^2, donde los pesos son los grados de libertad.

Intervalo de confianza para $\mu_1 - \mu_2$, $\sigma_1^2 = \sigma_2^2$ cuando se desconocen ambas varianzas

Si $\bar{x}_1$ y $\bar{x}_2$ son las medias de muestras aleatorias independientes con tamaños n_1 y n_2, respectivamente, tomadas de poblaciones más o menos normales con varianzas iguales pero desconocidas, un intervalo de confianza del $100(1-\alpha)\%$ para $\mu_1 - \mu_2$ es dado por

$$(\bar{x}_1 - \bar{x}_2) - t_{\alpha/2}s_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}} < \mu_1 - \mu_2 < (\bar{x}_1 - \bar{x}_2) + t_{\alpha/2}s_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}},$$

donde s_p es la estimación agrupada de la desviación estándar de la población y $t_{\alpha/2}$ es el valor t con $v = n_1 + n_2 - 2$ grados de libertad, que deja una área de $\alpha/2$ a la derecha.

Ejemplo 9.11: En el artículo “Estructura comunitaria de los macroinvertebrados como un indicador de la contaminación de minas ácidas”, publicado en el *Journal of Environmental Pollution*, se informa sobre una investigación realizada en Cane Creek, Alabama, para determinar la relación entre parámetros fisioquímicos seleccionados y diversas mediciones de la estructura de la comunidad de macroinvertebrados. Una faceta de la investigación consistió en evaluar la efectividad de un índice numérico de la diversidad de especies para indicar la degradación del agua debida al desagüe ácido de una mina. Conceptualmente, un índice elevado de la diversidad de especies macroinvertebradas debería indicar un sistema acuático no contaminado; mientras que un índice bajo de esta diversidad indicaría un sistema acuático contaminado.

Se eligieron 2 estaciones de muestreo independientes para este estudio: una que se localiza corriente abajo del punto de descarga ácida de la mina y la otra ubicada corriente arriba. Para 12 muestras mensuales reunidas en la estación corriente abajo el índice de diversidad de especies tuvo un valor medio de $\bar{x}_1 = 3.11$ y una desviación estándar de $s_1 = 0.771$; mientras que 10 muestras reunidas mensualmente en la estación corriente arriba tuvieron un valor medio del índice $\bar{x}_2 = 2.04$ y una desviación estándar de $s_2 = 0.448$. Calculemos un intervalo de confianza del 90% para la diferencia entre las medias de la población de los dos sitios, suponiendo que las poblaciones se distribuyen de forma aproximadamente normal y que tienen varianzas iguales.

***Solución*:** Representemos con μ_1 y μ_2 las medias de la población para los índices de diversidad de especies en las estaciones corriente abajo y corriente arriba, respectivamente. Deseamos encontrar un intervalo de confianza del 90% para $\mu_1 - \mu_2$. La estimación puntual de $\mu_1 - \mu_2$ es

$$\bar{x}_1 - \bar{x}_2 = 3.11 - 2.04 = 1.07$$

El estimado agrupado, s_p^2, de la varianza común, σ^2, es

$$s_p^2 = \frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2} = \frac{(11)(0.771^2) + (9)(0.448^2)}{12+10-2} = 0.417.$$

Al sacar la raíz cuadrada obtenemos $s_p = 0.646$. Si usamos $\alpha = 0.1$, encontramos en la tabla A.4 que $t_{0.05} = 1.725$ para $v = n_1 + n_2 - 2 = 20$ grados de libertad. Por lo tanto, el intervalo de confianza del 90% para $\mu_1 - \mu_2$ es

$$1.07 - (1.725)(0.646)\sqrt{\frac{1}{12}+\frac{1}{10}} < \mu_1 - \mu_2 < 1.07 + (1.725)(0.646)\sqrt{\frac{1}{12}+\frac{1}{10}},$$

que se simplifica a $0.593 < \mu_1 - \mu_2 < 1.547$.

Interpretación del intervalo de confianza

Para el caso de un solo parámetro el intervalo de confianza simplemente produce límites de error del parámetro. Los valores contenidos en el intervalo se deberían ver como valores razonables, dados los datos experimentales. En el caso de una diferencia entre dos medias, la interpretación se puede extender a una comparación de las dos medias. Por ejemplo, si tenemos gran confianza en que una diferencia $\mu_1 - \mu_2$ es positiva, sin duda inferiremos que $\mu_1 > \mu_2$ con poco riesgo de incurrir en un error. Así, en el ejemplo 9.11 tenemos un 90% de confianza en que el intervalo de 0.593 a 1.547 contiene la diferencia de las medias de la población para valores del índice de diversidad de especies en las dos estaciones. El hecho de que ambos límites de confianza sean positivos indica que, en promedio, el índice para la estación que se localiza corriente abajo del punto de descarga es mayor que el índice para la estación que se localiza corriente arriba.

Muestras de tamaños iguales

El procedimiento para construir intervalos de confianza para $\mu_1 - \mu_2$ cuando $\sigma_1 = \sigma_2 = \sigma$ pero ésta se desconoce, requiere suponer que las poblaciones son normales. Desviaciones ligeras de la suposición de varianzas iguales o de normalidad no alteran seriamente el grado de confianza en nuestro intervalo. (En el capítulo 10 se estudia un procedimiento para probar la igualdad de dos varianzas poblacionales desconocidas con base en la información que proporcionan las varianzas muestrales). Si las varianzas de la población son considerablemente diferentes, aún obtenemos resultados razonables cuando las poblaciones son normales, siempre y cuando $n_1 = n_2$. Por lo tanto, al planear un experimento se debería hacer un esfuerzo por igualar el tamaño de las muestras.

Varianzas desconocidas y distintas

Consideremos ahora el problema de calcular el estimado de un intervalo de $\mu_1 - \mu_2$ cuando no es probable que las varianzas de la población desconocidas sean iguales. El estadístico que se utiliza con mayor frecuencia en este caso es

$$T' = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{(S_1^2/n_1) + (S_2^2/n_2)}},$$

que tiene aproximadamente una distribución t con v grados de libertad, donde

$$v = \frac{(s_1^2/n_1 + s_2^2/n_2)^2}{[(s_1^2/n_1)^2/(n_1 - 1)] + [(s_2^2/n_2)^2/(n_2 - 1)]}.$$

Como v rara vez es un entero, lo *redondeamos* al número entero menor más cercano. El estimado anterior de los grados de libertad se denomina aproximación de Satterthwaite (Satterthwaite, 1946, en la bibliografía).

Con el estadístico T', escribimos

$$P(-t_{\alpha/2} < T' < t_{\alpha/2}) \approx 1 - \alpha,$$

donde $t_{\alpha/2}$ es el valor de la distribución t con v grados de libertad por arriba del cual encontramos una área de $\alpha/2$. Al sustituir para T' en la desigualdad y seguir los mismos pasos que antes, establecemos el resultado final.

Intervalo de confianza para $\mu_1 - \mu_2$, $\sigma_1^2 \neq \sigma_2^2$ y ambas varianzas se desconocen

Si $\bar{x}_1$ y s_1^2 y $\bar{x}_2$ y s_2^2 son las medias y varianzas de muestras aleatorias independientes de tamaños n_1 y n_2, respectivamente, tomadas de poblaciones aproximadamente normales con varianzas desconocidas y diferentes, un intervalo de confianza aproximado del 100(1 $- \alpha$)% para $\mu_1 - \mu_2$ es dado por

$$(\bar{x}_1 - \bar{x}_2) - t_{\alpha/2}\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} < \mu_1 - \mu_2 < (\bar{x}_1 - \bar{x}_2) + t_{\alpha/2}\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}},$$

donde $t_{\alpha/2}$ es el valor t con

$$v = \frac{(s_1^2/n_1 + s_2^2/n_2)^2}{[(s_1^2/n_1)^2/(n_1 - 1)] + [(s_2^2/n_2)^2/(n_2 - 1)]}$$

grados de libertad, que deja una área de $\alpha/2$ a la derecha.

Observe que la expresión para el valor v anterior incluye variables aleatorias y, por consiguiente, v es un *estimado* de los grados de libertad. En las aplicaciones este estimado no será un número entero, de manera que el analista lo debe redondear al entero menor más cercano para lograr la confianza que se busca.

Antes de ilustrar el intervalo de confianza anterior con un ejemplo deberíamos señalar que todos los intervalos de confianza para $\mu_1 - \mu_2$ tienen la misma forma general, como los de una sola media; a saber, se pueden escribir como

$$\text{estimación puntual} \pm t_{\alpha/2}\,\widehat{\text{e.e.}}(\text{estimación puntual})$$

o

$$\text{estimación puntual} \pm z_{\alpha/2}\,\text{e.e.}(\text{estimación puntual}).$$

Por ejemplo, en el caso donde $\sigma_1 = \sigma_2 = \sigma$, el error estándar estimado de $\bar{x}_1 - \bar{x}_2$ es $s_p\sqrt{1/n_1 + 1/n_2}$. Para el caso donde $\sigma_1^2 \neq \sigma_2^2$,

$$\widehat{\text{e.e.}}(\bar{x}_1 - \bar{x}_2) = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}.$$

Ejemplo 9.12: El Departamento de zoología de Virginia Tech llevó a cabo un estudio para estimar la diferencia en la cantidad de ortofósforo químico medido en dos estaciones diferentes del río James. El ortofósforo se mide en miligramos por litro. Se reunieron 15 muestras de la estación 1 y 12 muestras de la estación 2. Las 15 muestras de la estación 1 tuvieron un contenido promedio de ortofósforo de 3.84 miligramos por litro y una desviación estándar de 3.07 miligramos por litro; en tanto que las 12 muestras de la estación 2 tuvieron un contenido promedio de 1.49 miligramos por litro y una desviación estándar de 0.80 miligramos por litro. Calcule un intervalo de confianza de 95% para la diferencia en el contenido promedio verdadero de ortofósforo en estas dos estaciones. Suponga que las observaciones provienen de poblaciones normales con varianzas diferentes.

***Solución*:** Para la estación 1 tenemos $\bar{x}_1 = 3.84$, $s_1 = 3.07$ y $n_1 = 15$. Para la estación 2, $\bar{x}_2 = 1.49$, $s_2 = 0.80$ y $n_2 = 12$. Queremos obtener un intervalo de confianza del 95% para $\mu_1 - \mu_2$.

Como se suponen varianzas de la población diferentes, sólo podemos calcular un intervalo de confianza aproximado del 95% basado en la distribución t con v grados de libertad, donde

$$v = \frac{(3.07^2/15 + 0.80^2/12)^2}{[(3.07^2/15)^2/14] + [(0.80^2/12)^2/11]} = 16.3 \approx 16.$$

Nuestra estimación puntual de $\mu_1 - \mu_2$ es

$$\bar{x}_1 - \bar{x}_2 = 3.84 - 1.49 = 2.35.$$

Si usamos $\alpha = 0.05$, en la tabla A.4 encontramos que $t_{0.025} = 2.120$ para $v = 16$ grados de libertad. Por lo tanto, el intervalo de confianza del 95% para $\mu_1 - \mu_2$ es

$$2.35 - 2.120\sqrt{\frac{3.07^2}{15} + \frac{0.80^2}{12}} < \mu_1 - \mu_2 < 2.35 + 2.120\sqrt{\frac{3.07^2}{15} + \frac{0.80^2}{12}},$$

que se simplifica a $0.60 < \mu_1 - \mu_2 < 4.10$. En consecuencia, tenemos un 95% de confianza en que el intervalo de 0.60 a 4.10 miligramos por litro contiene la diferencia del promedio verdadero del ortofósforo que contienen estos dos lugares. ▟

Cuando se desconocen dos varianzas de la población, la suposición de varianzas iguales o diferentes podría ser precaria. En la sección 10.10 se presentará un procedimiento que ayudará a distinguir entre las situaciones con la misma varianza y con varianza diferente.

9.9 Observaciones pareadas

Ahora estudiaremos los procedimientos de estimación para la diferencia de dos medias cuando las muestras no son independientes y las varianzas de las dos poblaciones no son necesariamente iguales. La situación que se considera aquí tiene que ver con una condición experimental muy especial, a saber, *las observaciones pareadas*. A diferencia de la situación que se describió antes, las condiciones de las dos poblaciones no se asignan de forma aleatoria a las unidades experimentales. Más bien, cada unidad experimental homogénea recibe ambas condiciones de la población; como resultado, cada unidad experimental tiene un par de observaciones, una para cada población. Por ejemplo, si realizamos una prueba de una nueva dieta con 15 individuos, los pesos antes y después de seguir la dieta conforman la información de las dos muestras. Las dos poblaciones son "antes" y "después", y la unidad experimental es el individuo. Evidentemente, las observaciones en un par tienen algo en común. Para determinar si la dieta es efectiva consideramos las diferencias $d_1, d_2, \ldots, d_n$ en las observaciones pareadas. Estas diferencias son los valores de una muestra aleatoria $D_1, D_2, \ldots, D_n$ de una población de diferencias, que supondremos distribuidas normalmente, con media $\mu_D = \mu_1 - \mu_2$ y varianza σ_D^2. Estimamos σ_D^2 mediante s_d^2, la varianza de las diferencias que constituyen nuestra muestra. El estimador puntual de μ_D es dado por $\bar{D}$.

¿Cuándo debe hacerse el pareado?

Parear observaciones en un experimento es una estrategia que se puede emplear en muchos campos de aplicación. Se expondrá al lector a tal concepto en el material relacionado con la prueba de hipótesis en el capítulo 10 y en los temas de diseño experimental en los capítulos 13 y 15. Al seleccionar unidades experimentales relativamente homogéneas (dentro de las unidades) y permitir que cada unidad experimente ambas condiciones de la población, se reduce la varianza del error experimental efectiva (en este caso σ_D^2). El lector puede visualizar la i-ésima diferencia del par como

$$D_i = X_{1i} - X_{2i}.$$

Como las dos observaciones se toman de la unidad experimental de la muestra no son independientes y, de hecho,

$$\text{Var}(D_i) = \text{Var}(X_{1i} - X_{2i}) = \sigma_1^2 + \sigma_2^2 - 2\,\text{Cov}(X_{1i}, X_{2i}).$$

Entonces, de manera intuitiva, se espera que σ_D^2 debería reducirse debido a la similitud en la naturaleza de los "errores" de las dos observaciones dentro de una unidad experimental, a lo cual se llega mediante la expresión anterior. En realidad se espera que, si la unidad es homogénea, la covarianza sea positiva. Como resultado, la ganancia en calidad del intervalo de confianza sobre la que se obtuvo sin parear es mayor cuando hay homogeneidad dentro de las unidades y cuando las diferencias grandes van de una a otra unidad. Se debería tener en cuenta que el desempeño del intervalo de confianza dependerá del error estándar de $\bar{D}$, que es, por supuesto, $\sigma_D/\sqrt{n}$, donde n es el número de pares. Como indicamos antes, la intención al parear es reducir σ_D.

Equilibrio entre reducir la varianza y perder grados de libertad

Al comparar los intervalos de confianza obtenidos con y sin pareado es evidente que hay un intercambio implicado. Aunque en realidad el pareado debería reducir la varianza y, por lo tanto, el error estándar de la estimación puntual, los grados de libertad disminuyen al reducir el problema a uno con una sola muestra. Como resultado, el punto $t_{\alpha/2}$ ligado al error estándar se ajusta en concordancia. De esta manera, el pareado podría resultar contraproducente. Esto ocurriría con certeza si se experimenta sólo una reducción modesta en la varianza (a través de σ_D^2) mediante el pareado.

Otra ilustración del pareado implicaría elegir n pares de sujetos, donde cada par tenga una característica similar, como el coeficiente intelectual (CI), la edad o la raza, y luego para cada par seleccionar un miembro al azar para obtener un valor de X_1, dejando que el otro miembro proporcione el valor de X_2. En este caso, X_1 y X_2 podrían representar las calificaciones obtenidas por dos individuos con igual CI cuando uno es asignado al azar a un grupo que usa el método de enseñanza convencional y al otro a un grupo que utiliza materiales programados.

Se puede establecer un intervalo de confianza del 100(1 – α)% para μ_D escribiendo

$$P(-t_{\alpha/2} < T < t_{\alpha/2}) = 1 - \alpha,$$

donde $T = \frac{\bar{D} - \mu_D}{S_d/\sqrt{n}}$ y $t_{\alpha/2}$, como antes, es un valor de la distribución t con $n - 1$ grados de libertad.

En la actualidad se acostumbra reemplazar T por su definición en la desigualdad anterior y desarrollar los pasos matemáticos que conduzcan al siguiente intervalo de confianza del 100(1 – α)% para $\mu_1 - \mu_2 = \mu_D$.

Intervalo de confianza para $\mu_D = \mu_1 - \mu_2$ para observaciones pareadas

Si $\bar{d}$ y s_d son la media y la desviación estándar, respectivamente, de las diferencias distribuidas normalmente de n pares aleatorios de mediciones, un intervalo de confianza del $100(1-\alpha)\%$ para $\mu_D = \mu_1 - \mu_2$ es

$$\bar{d} - t_{\alpha/2}\frac{s_d}{\sqrt{n}} < \mu_D < \bar{d} + t_{\alpha/2}\frac{s_d}{\sqrt{n}},$$

donde $t_{\alpha/2}$ es el valor t con $v = n - 1$ grados de libertad, que deja una área de $\alpha/2$ a la derecha.

Ejemplo 9.13: Un estudio publicado en *Chemosphere* reporta los niveles de la dioxina TCDD en 20 veteranos de Vietnam de Massachusetts, quienes posiblemente estuvieron expuestos al agente naranja. En la tabla 9.1 se presentan los niveles de TCDD en plasma y tejido adiposo.

Calcule un intervalo de confianza del 95% para $\mu_1 - \mu_2$, donde μ_1 y μ_2 representen las medias verdaderas de los niveles de TCDD en plasma y en tejido adiposo, respectivamente. Suponga que la distribución de las diferencias es casi normal.

Tabla 9.1: Datos para el ejemplo 9.13.

Veterano	Niveles de TCDD en plasma	Niveles de TCDD en tejido adiposo	d_i	Veterano	Niveles de TCDD en plasma	Niveles de TCDD en tejido adiposo	d_i
1	2.5	4.9	−2.4	11	6.9	7.0	−0.1
2	3.1	5.9	−2.8	12	3.3	2.9	0.4
3	2.1	4.4	−2.3	13	4.6	4.6	0.0
4	3.5	6.9	−3.4	14	1.6	1.4	0.2
5	3.1	7.0	−3.9	15	7.2	7.7	−0.5
6	1.8	4.2	−2.4	16	1.8	1.1	0.7
7	6.0	10.0	−4.0	17	20.0	11.0	9.0
8	3.0	5.5	−2.5	18	2.0	2.5	−0.5
9	36.0	41.0	−5.0	19	2.5	2.3	0.2
10	4.7	4.4	0.3	20	4.1	2.5	1.6

Reproducido de *Chemosphere*, Vol. 20, Núms. 7-9 (tablas I y II), Schecter *et al.*, "Partitioning 2, 3, 7, 8-chlorinated dibenzo-p-dioxins and dibenzofurans between adipose tissue and plasma lipid of 20 Massachusetts Vietnam veterans", pp. 954-955, Derechos reservados ©1990, con autorización de Elsevier.

***Solución*:** Buscamos un intervalo de confianza del 95% para $\mu_1 - \mu_2$. Como las observaciones están pareadas, $\mu_1 - \mu_2 = \mu_D$. La estimación puntual de μ_D es $\bar{d} = -0.87$. La desviación estándar s_d de las diferencias muestrales es

$$s_d = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(d_i - \bar{d})^2} = \sqrt{\frac{168.4220}{19}} = 2.9773.$$

Si usamos $\alpha = 0.05$, en la tabla A.4 encontramos que $t_{0.025} = 2.093$ para $v = n - 1 = 19$ grados de libertad. Por lo tanto, el intervalo de confianza del 95% es

$$-0.8700 - (2.093)\left(\frac{2.9773}{\sqrt{20}}\right) < \mu_D < -0.8700 + (2.093)\left(\frac{2.9773}{\sqrt{20}}\right),$$

o simplemente $-2.2634 < \mu_D < 0.5234$, de lo cual concluimos que no hay diferencia significativa entre el nivel medio de TCDD en plasma y el nivel medio de TCDD en tejido adiposo.

Ejercicios

9.35 Una muestra aleatoria de tamaño $n_1 = 25$, tomada de una población normal con una desviación estándar $\sigma_1 = 5$, tiene una media $\bar{x}_1 = 80$. Una segunda muestra aleatoria de tamaño $n_2 = 36$, que se toma de una población normal diferente con una desviación estándar $\sigma_2 = 3$, tiene una media $\bar{x}_2 = 75$. Calcule un intervalo de confianza del 94% para $\mu_1 - \mu_2$.

9.36 Se comparan las resistencias de dos clases de hilo. Se prueban 50 piezas de cada clase de hilo en condiciones similares. La marca *A* tiene una resistencia a la tensión promedio de 78.3 kilogramos, con una desviación estándar de 5.6 kilogramos; en tanto que la marca *B* tiene una resistencia a la tensión promedio de 87.2 kilogramos con una desviación estándar de 6.3 kilogramos. Construya un intervalo de confianza del 95% para la diferencia de las medias de la población.

9.37 Se realiza un estudio para determinar si cierto tratamiento tiene algún efecto sobre la cantidad de metal que se elimina en una operación de encurtido. Una muestra aleatoria de 100 piezas se sumerge en un baño por 24 horas sin el tratamiento, lo que produce un promedio de 12.2 milímetros de metal eliminados y una desviación estándar muestral de 1.1 milímetros. Una segunda muestra de 200 piezas se somete al tratamiento, seguido de 24 horas de inmersión en el baño, lo que da como resultado una eliminación promedio de 9.1 milímetros de metal, con una desviación estándar muestral de 0.9 milímetros. Calcule un estimado del intervalo de confianza del 98% para la diferencia entre las medias de las poblaciones. ¿El tratamiento parece reducir la cantidad media del metal eliminado?

9.38 En un proceso químico por lotes se comparan los efectos de dos catalizadores sobre la potencia de la reacción del proceso. Se prepara una muestra de 12 lotes utilizando el catalizador 1 y una muestra de 10 lotes utilizando el catalizador 2. Los 12 lotes para los que se utilizó el catalizador 1 en la reacción dieron un rendimiento promedio de 85 con una desviación estándar muestral de 4; en tanto que para la segunda muestra, la de 10 lotes, el promedio fue de 81, con una desviación estándar muestral de 5. Calcule un intervalo de confianza del 90% para la diferencia entre las medias de la población, suponiendo que las poblaciones se distribuyen de forma aproximadamente normal y que tienen varianzas iguales.

9.39 Los estudiantes pueden elegir entre un curso de física de tres semestres-hora sin laboratorio y un curso de cuatro semestres-hora con laboratorio. El examen final escrito es el mismo para ambos cursos. Si 12 estudiantes del curso con laboratorio obtienen una calificación promedio de 84, con una desviación estándar de 4, y 18 estudiantes del grupo sin laboratorio obtienen una calificación promedio de 77, con una desviación estándar de 6, calcule un intervalo de confianza del 99% para la diferencia entre las calificaciones promedio para ambos cursos. Suponga que las poblaciones se distribuyen de forma aproximadamente normal y que tienen varianzas iguales.

9.40 En un estudio que se lleva a cabo en Virginia Tech sobre el desarrollo de micorriza, una relación simbiótica entre las raíces de árboles y un hongo, en la cual se transfieren minerales del hongo a los árboles y azúcares de los árboles a los hongos, se cultivaron en un invernadero 20 robles rojos que fueron expuestos al hongo *Pisolithus tinctorus.* Todos los árboles se plantaron en el mismo tipo de suelo y recibieron la misma cantidad de luz solar y agua. La mitad no recibió nitrógeno en el momento de plantarlos y sirvió como control, y la otra mitad recibió 368 ppm de nitrógeno en forma de $NaNO_3$. Después de 140 días se registraron los siguientes pesos de los tallos, en gramos:

Sin nitrógeno	Con nitrógeno
0.32	0.26
0.53	0.43
0.28	0.47
0.37	0.49
0.47	0.52
0.43	0.75
0.36	0.79
0.42	0.86
0.38	0.62
0.43	0.46

Construya un intervalo de confianza del 95% para la diferencia entre los pesos medios de los tallos que no recibieron nitrógeno y los que recibieron 368 ppm de nitrógeno. Suponga que las poblaciones están distribuidas normalmente y que tienen varianzas iguales.

9.41 Los siguientes datos representan el tiempo, en días, que pacientes tratados al azar con uno de dos medicamentos para curar infecciones graves de la vejiga tardaron en recuperarse:

Medicamento 1	Medicamento 2
$n_1 = 14$	$n_2 = 16$
$\bar{x}_1 = 17$	$\bar{x}_2 = 19$
$s_1^2 = 1.5$	$s_2^2 = 1.8$

Calcule un intervalo de confianza del 99% para la diferencia $\mu_2 - \mu_1$ en los tiempos medios de recuperación para los dos medicamentos. Suponga poblaciones normales que tienen varianzas iguales.

9.42 Un experimento publicado en *Popular Science* comparó el ahorro de combustible para dos tipos de camiones compactos que funcionan con diesel y están equipados de forma similar. Suponga que se utilizaron 12 camiones Volkswagen y 10 Toyota en pruebas con una velocidad constante de 90 kilómetros por hora. Si los 12 camiones Volkswagen promedian 16 kilómetros por litro con una desviación estándar de 1.0 kilómetros por litro, y los 10 Toyota promedian 11 kilómetros por litro con una desviación estándar de 0.8 kilómetros por litro, construya un intervalo de confianza del 90% para la diferencia entre los kilómetros promedio por litro de estos dos camiones compactos. Suponga que las distancias por litro para cada modelo de camión están distribuidas de forma aproximadamente normal y que tienen varianzas iguales.

9.43 Una empresa de taxis trata de decidir si comprará neumáticos de la marca *A* o de la marca *B* para su flotilla de taxis. Para estimar la diferencia entre las dos marcas realiza un experimento utilizando 12 neumáticos de cada marca, los cuales utiliza hasta que se desgastan. Los resultados son:

Marca *A*: $\bar{x}_1 = 36{,}300$ kilómetros,
$s_1 = 5000$ kilómetros.
Marca *B*: $\bar{x}_2 = 38{,}100$ kilómetros,
$s_2 = 6100$ kilómetros.

Calcule un intervalo de confianza del 95% para $\mu_A - \mu_B$, suponiendo que las poblaciones se distribuyen de forma aproximadamente normal. Puede no suponer que las varianzas son iguales.

9.44 Con referencia al ejercicio 9.43, calcule un intervalo de confianza del 99% para $\mu_1 - \mu_2$ si se asignan al azar neumáticos de las dos marcas a las ruedas traseras izquierda y derecha de 8 taxis y se registran las siguientes distancias, en kilómetros:

Taxi	Marca *A*	Marca *B*
1	34,400	36,700
2	45,500	46,800
3	36,700	37,700
4	32,000	31,100
5	48,400	47,800
6	32,800	36,400
7	38,100	38,900
8	30,100	31,500

Suponga que las diferencias de las distancias se distribuyen de forma aproximadamente normal.

9.45 El gobierno otorgó fondos para los departamentos de agricultura de 9 universidades para probar las capacidades de cosecha de dos nuevas variedades de trigo. Cada variedad se siembra en parcelas con la misma área en cada universidad, y las cosechas, en kilogramos por parcela, se registran como sigue:

	Universidad								
Variedad	1	2	3	4	5	6	7	8	9
1	38	23	35	41	44	29	37	31	38
2	45	25	31	38	50	33	36	40	43

Calcule un intervalo de confianza del 95% para la diferencia media entre las cosechas de las dos variedades, suponiendo que las diferencias entre las cosechas se distribuyen de forma aproximadamente normal. Explique por qué es necesario el pareado en este problema.

9.46 Los siguientes datos representan el tiempo de duración de películas producidas por dos empresas cinematográficas.

Empresa	Tiempo (minutos)						
I	103	94	110	87	98		
II	97	82	123	92	175	88	118

Calcule un intervalo de confianza del 90% para la diferencia entre la duración promedio de las películas que producen las dos empresas. Suponga que las diferencias en la duración se distribuyen de forma aproximadamente normal y que tienen varianzas distintas.

9.47 La revista *Fortune* (marzo de 1997) publicó la rentabilidad total de los inversionistas durante los 10 años anteriores a 1996 y también la de 431 empresas en ese mismo año. A continuación se lista la rentabilidad total para 10 de las empresas. Calcule un intervalo de confianza del 95% para el cambio promedio en el porcentaje de rentabilidad de los inversionistas.

	Rentabilidad total para los inversionistas	
Empresa	**1986–96**	**1996**
Coca-Cola	29.8%	43.3%
Mirage Resorts	27.9%	25.4%
Merck	22.1%	24.0%
Microsoft	44.5%	88.3%
Johnson & Johnson	22.2%	18.1%
Intel	43.8%	131.2%
Pfizer	21.7%	34.0%
Procter & Gamble	21.9%	32.1%
Berkshire Hathaway	28.3%	6.2%
S&P 500	11.8%	20.3%

9.48 Una empresa automotriz está considerando dos tipos de baterías para sus vehículos. Con ese fin reúne información muestral sobre la vida de las baterías. Utiliza para ello 20 baterías del tipo *A* y 20 baterías del tipo *B*. El resumen de los estadísticos es $\bar{x}_A = 32.91$,

$\bar{x}_B = 30.47$, $s_A = 1.57$ y $s_B = 1.74$. Suponga que los datos de cada batería se distribuyen normalmente y que $\sigma_A = \sigma_B$.

a) Calcule un intervalo de confianza del 95% para $\mu_A - \mu_B$.

b) Del inciso *a*) saque algunas conclusiones que le ayuden a la empresa a decidir si debería utilizar la batería *A* o la *B*.

9.49 Se considera usar dos marcas diferentes de pintura vinílica. Se seleccionaron 15 especímenes de cada tipo de pintura, para los cuales los tiempos de secado en horas fueron los siguientes:

Pintura *A*					**Pintura *B***				
3.5	2.7	3.9	4.2	3.6	4.7	3.9	4.5	5.5	4.0
2.7	3.3	5.2	4.2	2.9	5.3	4.3	6.0	5.2	3.7
4.4	5.2	4.0	4.1	3.4	5.5	6.2	5.1	5.4	4.8

Suponga que el tiempo de secado se distribuye normalmente, con $\sigma_A = \sigma_B$. Calcule un intervalo de confianza del 95% de $\mu_B - \mu_A$, donde μ_A y μ_B son los tiempos medios de secado.

9.50 A dos grupos de ratas diabéticas se les suministran dos niveles de dosis de insulina (alto y bajo) para verificar la capacidad de fijación de esta hormona. Se obtuvieron los siguientes datos.

Dosis baja:	$n_1 = 8$	$\bar{x}_1 = 1.98$	$s_1 = 0.51$
Dosis alta:	$n_2 = 13$	$\bar{x}_2 = 1.30$	$s_2 = 0.35$

Suponga que las varianzas son iguales. Determine un intervalo de confianza del 95% para la diferencia en la capacidad promedio verdadera de fijación de la insulina entre las dos muestras.

9.10 Una sola muestra: estimación de una proporción

El estadístico $\hat{P} = X/n$, en donde X representa el número de éxitos en n ensayos, provee un estimador puntual de la proporción p en un experimento binomial. Por lo tanto, la proporción de la muestra $\hat{p} = x/n$ se utilizará como el estimador puntual del parámetro p.

Si no se espera que la proporción p desconocida esté demasiado cerca de 0 o de 1, se puede establecer un intervalo de confianza para p considerando la distribución muestral de $\hat{P}$. Si en cada ensayo binomial asignamos el valor 0 a un fracaso y el valor 1 a un éxito, el número de éxitos, x, se puede interpretar como la suma de n valores que consta sólo de ceros y unos, y $\hat{p}$ es sólo la media muestral de esos n valores. En consecuencia, por el teorema del límite central, para n suficientemente grande $\hat{P}$ está distribuida de forma casi normal con media

$$\mu_{\hat{P}} = E(\hat{P}) = E\left(\frac{X}{n}\right) = \frac{np}{n} = p$$

y varianza

$$\sigma^2_{\hat{P}} = \sigma^2_{X/n} = \frac{\sigma^2_X}{n^2} = \frac{npq}{n^2} = \frac{pq}{n}.$$

Por lo tanto, podemos afirmar que

$$P(-z_{\alpha/2} < Z < z_{\alpha/2}) = 1 - \alpha, \text{ con } Z = \frac{\hat{P} - p}{\sqrt{pq/n}},$$

y $z_{\alpha/2}$ es el valor por arriba del cual encontramos una área de $\alpha/2$ debajo de la curva normal estándar. Al sustituir para Z escribimos

$$P\left(-z_{\alpha/2} < \frac{\hat{P} - p}{\sqrt{pq/n}} < z_{\alpha/2}\right) = 1 - \alpha.$$

Cuando n es grande se introduce un error muy pequeño sustituyendo el estimado puntual $\hat{p} = x/n$ para la p debajo del signo de radical. Entonces podemos escribir

$$P\left(\hat{P} - z_{\alpha/2}\sqrt{\frac{\hat{p}\hat{q}}{n}} < p < \hat{P} + z_{\alpha/2}\sqrt{\frac{\hat{p}\hat{q}}{n}}\right) \approx 1 - \alpha.$$

Por otro lado, al resolver para p en la desigualdad cuadrática anterior,

$$-z_{\alpha/2} < \frac{\hat{P} - p}{\sqrt{pq/n}} < z_{\alpha/2},$$

obtenemos otra forma del intervalo de confianza para p con los siguientes límites:

$$\frac{\hat{p} + \frac{z_{\alpha/2}^2}{2n}}{1 + \frac{z_{\alpha/2}^2}{n}} \pm \frac{z_{\alpha/2}}{1 + \frac{z_{\alpha/2}^2}{n}} \sqrt{\frac{\hat{p}\hat{q}}{n} + \frac{z_{\alpha/2}^2}{4n^2}}.$$

Para una muestra aleatoria de tamaño n se calcula la proporción muestral $\hat{p} = x/n$ y se pueden obtener los siguientes intervalos de confianza aproximados del $100(1 - \alpha)\%$ para p.

Intervalos de confianza para p de una muestra grande

Si $\hat{p}$ es la proporción de éxitos en una muestra aleatoria de tamaño n, y $\hat{q} = 1 - \hat{p}$, un intervalo de confianza aproximado del $100(1 - \alpha)\%$ para el parámetro binomial p se obtiene por medio de (método 1)

$$\hat{p} - z_{\alpha/2}\sqrt{\frac{\hat{p}\hat{q}}{n}} < p < \hat{p} + z_{\alpha/2}\sqrt{\frac{\hat{p}\hat{q}}{n}}$$

o mediante (método 2)

$$\frac{\hat{p} + \frac{z_{\alpha/2}^2}{2n}}{1 + \frac{z_{\alpha/2}^2}{n}} - \frac{z_{\alpha/2}}{1 + \frac{z_{\alpha/2}^2}{n}} \sqrt{\frac{\hat{p}\hat{q}}{n} + \frac{z_{\alpha/2}^2}{4n^2}} < p < \frac{\hat{p} + \frac{z_{\alpha/2}^2}{2n}}{1 + \frac{z_{\alpha/2}^2}{n}} + \frac{z_{\alpha/2}}{1 + \frac{z_{\alpha/2}^2}{n}} \sqrt{\frac{\hat{p}\hat{q}}{n} + \frac{z_{\alpha/2}^2}{4n^2}},$$

donde $z_{\alpha/2}$ es el valor z que deja una área de $\alpha/2$ a la derecha.

Cuando n es pequeña y se cree que la proporción desconocida p se acerca a 0 o a 1, el procedimiento del intervalo de confianza que se establece aquí no es confiable y, por lo tanto, no se debería emplear. Para estar seguros se requiere que tanto $n\hat{p}$ como $n\hat{q}$ sean mayores que o iguales a 5. Los métodos para calcular un intervalo de confianza para el parámetro binomial p también se pueden aplicar cuando se está utilizando la distribución binomial con el fin de aproximar la distribución hipergeométrica; es decir, cuando n es pequeña respecto a N, como se ilustra en el ejemplo 9.14.

Observe que, aunque el método 2 produce resultados más precisos, su cálculo es más complicado, y la ventaja en precisión que brinda disminuye cuando el tamaño de la muestra es lo suficientemente grande. Debido a esto en la práctica es más común utilizar el método 1.

Ejemplo 9.14: En una muestra aleatoria de $n = 500$ familias que tienen televisores en la ciudad de Hamilton, Canadá, se encuentra que $x = 340$ están suscritas a HBO. Calcule un intervalo de confianza del 95% para la proporción real de familias que tienen televisores en esta ciudad y están suscritas a HBO.

***Solución*:** La estimación puntual de p es $\hat{p} = 340/500 = 0.68$. Si usamos la tabla A.3, encontramos que $z_{0.025} = 1.96$. Por lo tanto, si utilizamos el método 1, el intervalo de confianza del 95% para p es

$$0.68 - 1.96\sqrt{\frac{(0.68)(0.32)}{500}} < p < 0.68 + 1.96\sqrt{\frac{(0.68)(0.32)}{500}},$$

que se simplifica a $0.6391 < p < 0.7209$.

Si utilizamos el segundo método, obtenemos

$$\frac{0.68+\frac{1.96^2}{(2)(500)}}{1+\frac{1.96^2}{500}} \pm \frac{1.96}{1+\frac{1.96^2}{500}}\sqrt{\frac{(0.68)(0.32)}{500}+\frac{1.96^2}{(4)(500^2)}} = 0.6786 \pm 0.0408,$$

que se simplifica a $0.6378 < p < 0.7194$. Aparentemente, cuando n es grande (500 en este caso) ambos métodos producen resultados muy similares.

Si p es el valor central de un intervalo de confianza del $100(1-\alpha)\%$, entonces $\hat{p}$ estima p sin error. Sin embargo, la mayoría de las veces $\hat{p}$ no será exactamente igual a p y el estimado puntual será erróneo. El tamaño de este error será la diferencia positiva que separa a p de $\hat{p}$, y podemos tener una confianza del $100(1-\alpha)\%$ de que tal diferencia no excederá a $z_{\alpha/2}\sqrt{\hat{p}\hat{q}/n}$. Si dibujamos un diagrama de un intervalo de confianza típico, como el de la figura 9.6, podemos ver esto fácilmente. En este caso utilizamos el método 1 para estimar el error.

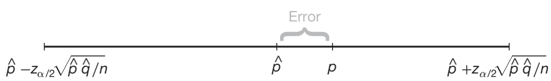

Figura 9.6: Error en la estimación de p por medio de $\hat{p}$.

Teorema 9.3: Si $\hat{p}$ se utiliza como un estimado de p, podemos tener un $100(1-\alpha)\%$ de confianza en que el error no excederá a $z_{\alpha/2}\sqrt{\hat{p}\hat{q}/n}$.

En el ejemplo 9.14 tenemos un 95% de confianza en que la proporción de la muestra $\hat{p} = 0.68$ difiere de la verdadera proporción p en una cantidad que no excede a 0.04.

Selección del tamaño de la muestra

Determinemos ahora qué tan grande debe ser una muestra para poder estar seguros de que el error al estimar p será menor que una cantidad específica e. Por medio del teorema 9.3, debemos elegir una n tal que $z_{\alpha/2}\sqrt{\hat{p}\hat{q}/n} = e$.

Teorema 9.4: Si $\hat{p}$ se utiliza como un estimado de p, podemos tener un $100(1-\alpha)\%$ de confianza en que el error será menor que una cantidad específica e cuando el tamaño de la muestra sea aproximadamente

$$n = \frac{z_{\alpha/2}^2\hat{p}\hat{q}}{e^2}.$$

El teorema 9.4 es algo engañoso, pues debemos utilizar $\hat{p}$ para determinar el tamaño n de la muestra, pero $\hat{p}$ se calcula a partir de la muestra. Si se puede hacer una estimación burda de p sin tomar una muestra, se podría usar este valor para determinar n. A falta de tal estimado, podríamos tomar una muestra preliminar de tamaño $n \geq 30$ para proporcionar un estimado de p. Si utilizamos el teorema 9.4 podríamos determinar aproximadamente cuántas observaciones se necesitan para proporcionar el grado de precisión deseado. Observe que los valores fraccionarios de n se redondean al siguiente número entero mayor.

Ejemplo 9.15: ¿Qué tan grande debe ser una muestra en el ejemplo 9.14 si queremos tener un 95% de confianza en que la estimación de p esté dentro de 0.02 del valor verdadero?

***Solución*:** Tratemos a las 500 familias como una muestra preliminar que proporciona una estimación $\hat{p} = 0.68$. Entonces, mediante el teorema 9.4,

$$n = \frac{(1.96)^2(0.68)(0.32)}{(0.02)^2} = 2089.8 \approx 2090.$$

Por lo tanto, si basamos nuestra estimación de p en una muestra aleatoria de tamaño 2090, podemos tener un 95% de confianza en que nuestra proporción muestral no diferirá de la proporción verdadera en más de 0.02.

Ocasionalmente será poco práctico obtener una estimación de p que se utilice para determinar el tamaño muestral para un grado específico de confianza. Si esto sucede, se establece un límite superior para n al notar que $\hat{p}\hat{q} = \hat{p}(1-\hat{p})$, que debe ser a lo sumo $1/4$, ya que $\hat{p}$ debe caer entre 0 y 1. Este hecho se verifica completando el cuadrado. Por consiguiente,

$$\hat{p}(1-\hat{p}) = -(\hat{p}^2 - \hat{p}) = \frac{1}{4} - \left(\hat{p}^2 - \hat{p} + \frac{1}{4}\right) = \frac{1}{4} - \left(\hat{p} - \frac{1}{2}\right)^2,$$

que siempre es menor que $1/4$ excepto cuando $\hat{p} = 1/2$ y entonces $\hat{p}\hat{q} = 1/4$. Por lo tanto, si sustituimos $\hat{p} = 1/2$ en la fórmula para n del teorema 9.4, cuando, de hecho, p difiere de $1/2$, entonces n se agrandará más de lo necesario para el grado de confianza específico y, como resultado, se incrementará nuestro grado de confianza.

Teorema 9.5: Si utilizamos $\hat{p}$ como un estimado de p, podemos tener, **al menos**, un $100(1-\alpha)\%$ de confianza en que el error no excederá a una cantidad específica e cuando el tamaño de la muestra sea

$$n = \frac{z_{\alpha/2}^2}{4e^2}.$$

Ejemplo 9.16: ¿Qué tan grande debe ser una muestra en el ejemplo 9.14 si queremos tener al menos un 95% de confianza en que nuestra estimación de p está dentro de 0.02 del valor verdadero?

***Solución*:** A diferencia del ejemplo 9.15, supondremos ahora que no se tomó una muestra preliminar para obtener una estimación de p. En consecuencia, podemos tener al menos un 95% de confianza en que nuestra proporción de la muestra no diferirá de la proporción verdadera en más de 0.02, si elegimos una muestra de tamaño

$$n = \frac{(1.96)^2}{(4)(0.02)^2} = 2401.$$

Si comparamos los resultados de los ejemplos 9.15 y 9.16, vemos que la información concerniente a p, proporcionada por una muestra preliminar, o quizás obtenida a partir de la experiencia, nos permite elegir una muestra más pequeña a la vez que mantenemos el grado de precisión requerido.

9.11 Dos muestras: estimación de la diferencia entre dos proporciones

Considere el problema en el que se busca estimar la diferencia entre dos parámetros binomiales p_1 y p_2. Por ejemplo, p_1 podría ser la proporción de fumadores con cáncer de pulmón y p_2 la proporción de no fumadores con cáncer de pulmón, y el problema consistiría en estimar la diferencia entre estas dos proporciones. Primero seleccionamos muestras aleatorias independientes de tamaños n_1 y n_2 a partir de las dos poblaciones binomiales con medias n_1p_1 y n_2p_2, y varianzas $n_1p_1q_1$ y $n_2p_2q_2$, respectivamente, después determinamos los números x_1 y x_2 de personas con cáncer de pulmón en cada muestra, y formamos las proporciones $\hat{p}_1 = x_1/n$ y $\hat{p}_2 = x_2/n$. El estadístico $\hat{P}_1 - \hat{P}_2$ provee un estimador puntual de la diferencia entre las dos proporciones, $p_1 - p_2$. Por lo tanto, la diferencia de las proporciones muestrales, $\hat{p}_1 - \hat{p}_2$, se utilizará como la estimación puntual de $p_1 - p_2$.

Se puede establecer un intervalo de confianza para $p_1 - p_2$ considerando la distribución muestral de $\hat{P}_1 - \hat{P}_2$. De la sección 9.10 sabemos que $\hat{P}_1$ y $\hat{P}_2$ están distribuidos cada uno de forma aproximadamente normal, con medias p_1 y p_2, y varianzas p_1q_1/n_1 y p_2q_2/n_2, respectivamente. Al elegir muestras independientes de las dos poblaciones nos aseguramos de que las variables $\hat{P}_1$ y $\hat{P}_2$ serán independientes y luego, por la propiedad reproductiva de la distribución normal que se estableció en el teorema 7.11, concluimos que $\hat{P}_1 - \hat{P}_2$ está distribuido de forma aproximadamente normal con media

$$\mu_{\hat{P}_1 - \hat{P}_2} = p_1 - p_2$$

y varianza

$$\sigma^2_{\hat{P}_1 - \hat{P}_2} = \frac{p_1q_1}{n_1} + \frac{p_2q_2}{n_2}.$$

Por lo tanto, podemos asegurar que

$$P(-z_{\alpha/2} < Z < z_{\alpha/2}) = 1 - \alpha,$$

donde

$$Z = \frac{(\hat{P}_1 - \hat{P}_2) - (p_1 - p_2)}{\sqrt{p_1q_1/n_1 + p_2q_2/n_2}}$$

y $z_{\alpha/2}$ es un valor por arriba del cual encontramos una área de $\alpha/2$ debajo de la curva normal estándar. Al sustituir para Z escribimos

$$P\left[-z_{\alpha/2} < \frac{(\hat{P}_1 - \hat{P}_2) - (p_1 - p_2)}{\sqrt{p_1q_1/n_1 + p_2q_2/n_2}} < z_{\alpha/2}\right] = 1 - \alpha.$$

Después de realizar las operaciones matemáticas usuales reemplazamos p_1, p_2, q_1 y q_2 bajo el signo de radical por sus estimaciones $\hat{p}_1 = x_1/n_1$, $\hat{p}_2 = x_2/n_2$, $\hat{q}_1 = 1 - \hat{p}_1$ y $\hat{q}_2 = 1 - \hat{p}_2$, siempre y cuando $n_1\hat{p}_1$, $n_2\hat{q}_1$, $n_2\hat{p}_2$ y $n_1\hat{q}_1$ sean todas mayores que o iguales a 5, y se obtiene el siguiente intervalo de confianza aproximado del $100(1 - \alpha)\%$ para $p_1 - p_2$.

Intervalo de confianza para $p_1 - p_2$ de una muestra grande

Si $\hat{p}_1$ y $\hat{p}_2$ son las proporciones de éxitos en muestras aleatorias de tamaños n_1 y n_2, respectivamente, $\hat{q}_1 = 1 - \hat{p}_1$ y $\hat{q}_2 = 1 - \hat{p}_2$, un intervalo de confianza aproximado del $100(1 - \alpha)\%$ para la diferencia de dos parámetros binomiales $p_1 - p_2$ es dado por

$$(\hat{p}_1 - \hat{p}_2) - z_{\alpha/2}\sqrt{\frac{\hat{p}_1\hat{q}_1}{n_1} + \frac{\hat{p}_2\hat{q}_2}{n_2}} < p_1 - p_2 < (\hat{p}_1 - \hat{p}_2) + z_{\alpha/2}\sqrt{\frac{\hat{p}_1\hat{q}_1}{n_1} + \frac{\hat{p}_2\hat{q}_2}{n_2}},$$

donde $z_{\alpha/2}$ es el valor z que deja una área de $\alpha/2$ a la derecha.

Ejemplo 9.17: Se considera hacer un cierto cambio en el proceso de fabricación de partes componentes. Para determinar si el cambio en el proceso da como resultado una mejora, se toman muestras de partes fabricadas con el proceso nuevo y con el actual. Si se encuentra que 75 de 1500 artículos manufacturados con el proceso actual están defectuosos y 80 de 2000 manufacturados con el proceso nuevo también lo están, calcule un intervalo de confianza del 90% para la diferencia verdadera en la proporción de partes defectuosas entre el proceso actual y el nuevo.

***Solución*:** Suponga que p_1 y p_2 son las proporciones verdaderas de partes defectuosas para los procesos actual y nuevo, respectivamente. En consecuencia, $\hat{p}_1 = 75/1500 = 0.05$ y $\hat{p}_2 = 80/2000 = 0.04$, y la estimación puntual de $p_1 - p_2$ es

$$\hat{p}_1 - \hat{p}_2 = 0.05 - 0.04 = 0.01.$$

Si utilizamos la tabla A.3, encontramos $z_{0.05} = 1.645$. Por lo tanto, al sustituir en la fórmula

$$1.645\sqrt{\frac{(0.05)(0.95)}{1500} + \frac{(0.04)(0.96)}{2000}} = 0.0117,$$

encontramos que el intervalo de confianza del 90% es $-0.0017 < p_1 - p_2 < 0.0217$. Como el intervalo contiene el valor 0, no hay razón para creer que el nuevo proceso, comparado con el actual, disminuye en forma significativa la proporción de artículos defectuosos.

Hasta aquí todos los intervalos de confianza presentados son de la forma

$$\text{estimación puntual} \pm K \text{ e.e.(estimación puntual)},$$

donde K es una constante (ya sea t o el punto porcentual normal). Esta forma es válida cuando el parámetro es una media, una diferencia entre medias, una proporción o una diferencia entre proporciones, debido a la simetría de las distribuciones t y Z. Sin embargo, no se extiende a las varianzas ni a los cocientes de las varianzas, las cuales se examinarán en las secciones 9.12 y 9.13.

Ejercicios

En este conjunto de ejercicios, para una estimación respecto a una proporción, utilice sólo el método 1 para calcular los intervalos de confianza, a menos que se especifique otra cosa.

9.51 En una muestra aleatoria de 1000 viviendas en cierta ciudad se encuentra que 228 utilizan petróleo como combustible para la calefacción. Calcule intervalos de confianza del 99% para la proporción de viviendas en esta ciudad que utilizan petróleo con el fin mencionado. Utilice los dos métodos que se presentaron en la página 297.

9.52 Calcule intervalos de confianza del 95% para la proporción de artículos defectuosos que resultan de un proceso cuando se encuentra que una muestra de tamaño 100 produce 8 defectuosos. Utilice los dos métodos que se presentaron en la página 297.

9.53 *a*) Se selecciona una muestra aleatoria de 200 votantes en una ciudad y se encuentra que 114 apoyan un juicio de anexión. Calcule el intervalo de confianza del 96% para la parte de la población votante que está a favor del juicio.

b) ¿Qué podemos afirmar con 96% de confianza acerca de la posible magnitud de nuestro error, si estimamos que la fracción de votantes que está a favor del juicio de anexión es 0.57?

9.54 Un fabricante de reproductores de MP3 utiliza un conjunto de pruebas exhaustivas para evaluar el funcionamiento eléctrico de su producto. Todos los reproductores de MP3 deben pasar todas las pruebas antes de ser puestos a la venta. De una muestra aleatoria de 500 reproductores, 15 no pasan una o más de las pruebas. Calcule un intervalo de confianza del 90% para la proporción de los reproductores de MP3 de la población que pasan todas las pruebas.

9.55 Se está considerando un nuevo sistema de lanzamiento de cohetes para el despliegue de cohetes pequeños, de corto alcance. La probabilidad de que el sistema existente tenga un lanzamiento exitoso se representa con $p = 0.8$. Se toma una muestra de 40 lanzamientos experimentales con el nuevo sistema y 34 resultan exitosos.

a) Construya un intervalo de confianza del 95% para p.

b) ¿Con base en sus resultados, concluiría que el nuevo sistema es mejor?

9.56 Un genetista está interesado en determinar la proporción de hombres africanos que padecen cierto trastorno sanguíneo menor. En una muestra aleatoria de 100 hombres africanos encuentra que 24 lo padecen.

a) Calcule un intervalo de confianza del 99% para la proporción de hombres africanos que padecen este trastorno sanguíneo.

b) ¿Qué podríamos afirmar con 99% de confianza acerca de la posible magnitud de nuestro error, si estimamos que la proporción de hombres africanos con dicho trastorno sanguíneo es 0.24?

9.57 *a*) De acuerdo con un reporte del *Roanoke Times & World-News*, aproximadamente $2/3$ de los 1600 adultos encuestados vía telefónica dijeron que piensan que invertir en el programa del transbordador espacial es bueno para Estados Unidos. Calcule un intervalo de confianza del 95% para la proporción de adultos estadounidenses que piensan que el programa del transbordador espacial es una buena inversión para su país.

b) ¿Qué podríamos afirmar con un 95% de confianza acerca de la posible magnitud de nuestro error, si estimamos que la proporción de adultos estadounidenses que piensan que el programa del transbordador espacial es una buena inversión es de $2/3$?

9.58 En el artículo del periódico al que se hace referencia en el ejercicio 9.57, 32% de los 1600 adultos encuestados dijo que el programa espacial estadounidense debería enfatizar la exploración científica. ¿Qué tamaño debería tener una muestra de adultos para la encuesta si se desea tener un 95% de confianza en que el porcentaje estimado esté dentro del 2% del porcentaje verdadero?

9.59 ¿Qué tamaño debería tener una muestra si deseamos tener un 96% de confianza en que nuestra proporción de la muestra en el ejercicio 9.53 esté dentro del 0.02 de la fracción verdadera de la población votante?

9.60 ¿Qué tamaño debería tener una muestra si deseamos tener un 99% de confianza en que nuestra proporción de la muestra en el ejercicio 9.51 esté dentro del 0.05 de la proporción verdadera de viviendas en esa ciudad que utilizan petróleo como combustible para la calefacción?

9.61 ¿Qué tamaño debería tener una muestra en el ejercicio 9.52 si deseamos tener un 98% de confianza en que nuestra proporción de la muestra esté dentro del 0.05 de la proporción verdadera de defectuosos?

9.62 Una conjetura de un catedrático del departamento de microbiología, de la Facultad de Odontología de la Universidad de Washington, en St. Louis, Missouri, afirma que un par de tasas diarias de té verde o negro proporciona suficiente flúor para evitar el deterioro de los dientes. ¿Qué tan grande debería ser la muestra para estimar el porcentaje de habitantes de cierta ciudad que están a favor de tener agua fluorada, si se desea tener al menos un 99% de confianza en que el estimado está dentro del 1% del porcentaje verdadero?

9.63 Se llevará a cabo un estudio para estimar el porcentaje de ciudadanos de una ciudad que están a favor de tener agua fluorada. ¿Qué tan grande debería ser la muestra si se desea tener al menos 95% de confianza en que el estimado esté dentro del 1% del porcentaje verdadero?

9.64 Se realizará un estudio para estimar la proporción de residentes de cierta ciudad y sus suburbios que está a favor de que se construya una planta de energía nuclear cerca de la ciudad. ¿Qué tan grande debería ser la muestra, si se desea tener al menos un 95% de confianza en que el estimado esté dentro del 0.04 de la verdadera proporción de residentes que están a favor de que se construya la planta de energía nuclear?

9.65 A cierto genetista le interesa determinar la proporción de hombres y mujeres de la población que padecen cierto trastorno sanguíneo menor. En una muestra aleatoria de 1000 hombres encuentra que 250 lo padecen; mientras que de 1000 mujeres examinadas, 275 parecen padecerlo. Calcule un intervalo de confianza del 95% para la diferencia entre la proporción de hombres y mujeres que padecen el trastorno sanguíneo.

9.66 Se encuestan 10 escuelas de ingeniería de Estados Unidos. La muestra contiene a 250 ingenieros eléctricos, de los cuales 80 son mujeres; y 175 ingenieros químicos, de los cuales 40 son mujeres. Calcule un intervalo de confianza del 90% para la diferencia entre la proporción de mujeres en estos dos campos de la ingeniería. ¿Hay una diferencia significativa entre las dos proporciones?

9.67 Se llevó a cabo una prueba clínica para determinar si cierto tipo de vacuna tiene un efecto sobre la incidencia de cierta enfermedad. Una muestra de 1000 ratas, 500 de las cuales recibieron la vacuna, se mantuvo en un ambiente controlado durante un periodo de un año. En el grupo que no fue vacunado, 120 ratas presentaron la enfermedad, mientras que en el grupo inoculado 98 ratas la contrajeron. Si p_1 es la probabilidad de incidencia de la enfermedad en las ratas sin vacuna y p_2 es la probabilidad de incidencia en las ratas inoculadas, calcule un intervalo de confianza del 90% para $p_1 - p_2$.

9.68 En el estudio *Germination and Emergence of Broccoli,* realizado por el Departamento de horticultura del Virginia Tech, un investigador encontró que a 5°C, de 20 semillas de brócoli germinaron 10; en tanto que a 15°C, de 20 semillas germinaron 15. Calcule un intervalo de confianza del 95% para la diferencia en la proporción de semillas que germinaron a las dos temperaturas y decida si esta diferencia es significativa.

9.69 Una encuesta de 1000 estudiantes reveló que 274 eligen al equipo profesional de beisbol *A* como su equipo favorito. En 1991 se realizó una encuesta similar con 760 estudiantes y 240 de ellos también eligieron a ese equipo como su favorito. Calcule un intervalo de confianza del 95% para la diferencia entre la proporción de estudiantes que favorecen al equipo *A* en las dos encuestas. ¿Hay una diferencia significativa?

9.70 De acuerdo con el *USA Today* (17 de marzo de 1997), las mujeres constituían el 33.7% del personal de redacción en las estaciones locales de televisión en 1990 y el 36.2% en 1994. Suponga que en 1990 y en 1994 se contrataron 20 nuevos empleados para el personal de redacción.

a) Estime el número de trabajadores que habrían sido mujeres en 1990 y en 1994, respectivamente.

b) Calcule un intervalo de confianza del 95% para saber si hay evidencia de que la proporción de mujeres contratadas para el equipo de redacción fue mayor en 1994 que en 1990.

9.12 Una sola muestra: estimación de la varianza

Si extraemos una muestra de tamaño n de una población normal con varianza σ^2 y calculamos la varianza muestral s^2, obtenemos un valor del estadístico S^2. Esta varianza muestral calculada se utiliza como una estimación puntual de σ^2. En consecuencia, al estadístico S^2 se le denomina estimador de σ^2

Se puede establecer una estimación por intervalos de σ^2 utilizando el estadístico

$$X^2 = \frac{(n-1)S^2}{\sigma^2}.$$

De acuerdo con el teorema 8.4, cuando las muestras se toman de una población normal el estadístico X^2 tiene una distribución chi cuadrada con $n - 1$ grados de libertad. Podemos escribir (véase la figura 9.7)

$$P(\chi^2_{1-\alpha/2} < X^2 < \chi^2_{\alpha/2}) = 1 - \alpha,$$

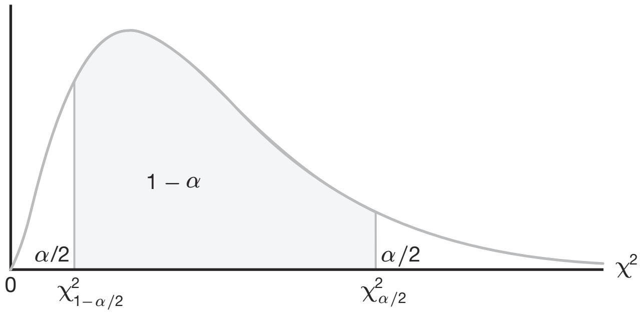

Figura 9.7: $P\left(\chi^2_{1-\alpha/2} < X^2 < \chi^2_{\alpha/2}\right) = 1 - \alpha$.

donde $\chi^2_{1-\alpha/2}$ y $\chi^2_{\alpha/2}$ son valores de la distribución chi cuadrada con $n - 1$ grados de libertad, que dejan áreas de $1 - \alpha/2$ y $\alpha/2$, respectivamente, a la derecha. Al sustituir para X^2 escribimos

$$P\left[\chi^2_{1-\alpha/2} < \frac{(n-1)S^2}{\sigma^2} < \chi^2_{\alpha/2}\right] = 1 - \alpha.$$

Si dividimos cada término de la desigualdad entre $(n - 1)S^2$, y después invertimos cada término (lo que cambia el sentido de las desigualdades), obtenemos

$$P\left[\frac{(n-1)S^2}{\chi^2_{\alpha/2}} < \sigma^2 < \frac{(n-1)S^2}{\chi^2_{1-\alpha/2}}\right] = 1 - \alpha.$$

Para una muestra aleatoria de tamaño n, tomada de una población normal, se calcula la varianza muestral s^2 y se obtiene el siguiente intervalo de confianza del $100(1 - \alpha)\%$ para σ^2.

Intervalo de confianza para σ^2

Si s^2 es la varianza de una muestra aleatoria de tamaño n de una población normal, un intervalo de confianza del $100(1 - \alpha)\%$ para σ^2 es

$$\frac{(n-1)s^2}{\chi^2_{\alpha/2}} < \sigma^2 < \frac{(n-1)s^2}{\chi^2_{1-\alpha/2}},$$

donde $\chi^2_{\alpha/2}$ y $\chi^2_{1-\alpha/2}$ son valores χ^2 con $v = n - 1$ grados de libertad, que dejan áreas de $\alpha/2$ y $1 - \alpha/2$, respectivamente, a la derecha.

Un intervalo de confianza aproximado a $100(1 - \alpha)\%$ para σ se obtiene tomando la raíz cuadrada de cada extremo del intervalo para σ^2.

Ejemplo 9.18: Los siguientes son los pesos, en decagramos, de 10 paquetes de semillas de pasto distribuidas por cierta empresa: 46.4, 46.1, 45.8, 47.0, 46.1. 45.9, 45.8, 46.9, 45.2 y 46.0. Calcule un intervalo de confianza del 95% para la varianza de todos los pesos de este tipo de paquetes de semillas de pasto distribuidos por la empresa. Suponga una población normal.

***Solución*:** Primero calculamos

$$s^2 = \frac{n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2}{n(n-1)} = \frac{(10)(21{,}273.12) - (461.2)^2}{(10)(9)} = 0.286.$$

Para obtener un intervalo de confianza del 95% elegimos $\alpha = 0.05$. Después, usando la tabla A.5 con $v = 9$ grados de libertad, encontramos $\chi^2_{0.025} = 19.023$ y $\chi^2_{0.975} = 2.700$. Por lo tanto, el intervalo de confianza del 95% para σ^2 es

$$\frac{(9)(0.286)}{19.023} < \sigma^2 < \frac{(9)(0.286)}{2.700},$$

o simplemente $0.135 < \sigma^2 < 0.953$.

9.13 Dos muestras: estimación de la proporción de dos varianzas

Una estimación puntual de la proporción de dos varianzas de la población σ_1^2/σ_2^2 es dada por la proporción s_1^2/s_2^2 de las varianzas muestrales. En consecuencia, el estadístico S_1^2/S_2^2 se conoce como un estimador de σ_1^2/σ_2^2.

Si σ_1^2 y σ_2^2 son las varianzas de poblaciones normales, podemos establecer una estimación por intervalos de σ_1^2/σ_2^2 usando el estadístico

$$F = \frac{\sigma_2^2 S_1^2}{\sigma_1^2 S_2^2}.$$

De acuerdo con el teorema 8.8, la variable aleatoria F tiene una distribución F con $v_1 = n_1 - 1$ y $v_2 = n_2 - 1$ grados de libertad. Por lo tanto, podemos escribir (véase la figura 9.8)

$$P[f_{1-\alpha/2}(v_1, v_2) < F < f_{\alpha/2}(v_1, v_2)] = 1 - \alpha,$$

donde $f_{1-\alpha/2}(v_1, v_2)$ y $f_{\alpha/2}(v_1, v_2)$ son los valores de la distribución F con v_1 y v_2 grados de libertad, que dejan áreas de $1 - \alpha/2$ y $\alpha/2$, respectivamente, a la derecha.

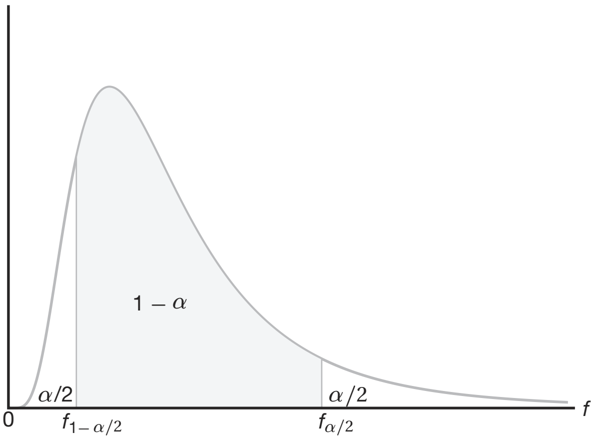

Figura 9.8: $P[f_{1-\alpha/2}(v_1, v_2) < F < f_{\alpha/2}(v_1, v_2)] = 1 - \alpha$.

Al sustituir para F, escribimos

$$P\left[f_{1-\alpha/2}(v_1, v_2) < \frac{\sigma_2^2 S_1^2}{\sigma_1^2 S_2^2} < f_{\alpha/2}(v_1, v_2)\right] = 1 - \alpha.$$

Si multiplicamos cada término de la desigualdad por S_2^2/S_1^2, y después invertimos cada término, obtenemos

$$P\left[\frac{S_1^2}{S_2^2}\frac{1}{f_{\alpha/2}(v_1, v_2)} < \frac{\sigma_1^2}{\sigma_2^2} < \frac{S_1^2}{S_2^2}\frac{1}{f_{1-\alpha/2}(v_1, v_2)}\right] = 1 - \alpha.$$

Los resultados del teorema 8.7 nos permiten reemplazar la cantidad $f_{1-\alpha/2}(v_1, v_2)$ por $1/f_{\alpha/2}(v_2, v_1)$. Por lo tanto,

$$P\left[\frac{S_1^2}{S_2^2}\frac{1}{f_{\alpha/2}(v_1, v_2)} < \frac{\sigma_1^2}{\sigma_2^2} < \frac{S_1^2}{S_2^2} f_{\alpha/2}(v_2, v_1)\right] = 1 - \alpha.$$

Para cualesquiera dos muestras aleatorias independientes de tamaño n_1 y n_2 que se seleccionan de dos poblaciones normales, se calcula la proporción de las varianzas muestrales s_1^2/s_2^2 y se obtiene el siguiente intervalo de confianza del $100(1-\alpha)\%$ para σ_1^2/σ_2^2.

Intervalo de confianza para σ_1^2/σ_2^2

Si s_1^2 y s_2^2 son las varianzas de muestras independientes de tamaño n_1 y n_2, respectivamente, tomadas de poblaciones normales, entonces un intervalo de confianza del $100(1-\alpha)\%$ para σ_1^2/σ_2^2 es

$$\frac{s_1^2}{s_2^2}\frac{1}{f_{\alpha/2}(v_1, v_2)} < \frac{\sigma_1^2}{\sigma_2^2} < \frac{s_1^2}{s_2^2} f_{\alpha/2}(v_2, v_1),$$

donde $f_{\alpha/2}(v_1, v_2)$ es un valor f con $v_1 = n_1 - 1$ y $v_2 = n_2 - 1$ grados de libertad que deja una área de $\alpha/2$ a la derecha, y $f_{\alpha/2}(v_2, v_1)$ es un valor f similar con $v_2 = n_2 - 1$ y $v_1 = n_1 - 1$ grados de libertad.

Como vimos en la sección 9.12, tomando la raíz cuadrada de cada extremo del intervalo para σ_1^2/σ_2^2, se obtiene un intervalo de confianza del $100(1-\alpha)\%$ para σ_1/σ_2.

Ejemplo 9.19: En el ejemplo 9.12 de la página 290 se construyó un intervalo de confianza para la diferencia en el contenido medio de ortofósforo de dos estaciones ubicadas sobre el río James, medido en miligramos por litro, suponiendo que las varianzas normales de la población son diferentes. Justifique esta suposición construyendo intervalos de confianza del 98% para σ_1^2/σ_2^2 y para σ_1/σ_2, donde σ_1^2 y σ_2^2 son las varianzas de la población del contenido de ortofósforo en la estación 1 y en la estación 2, respectivamente.

***Solución*:** Del ejemplo 9.12 tenemos $n_1 = 15$, $n_2 = 12$, $s_1 = 3.07$ y $s_2 = 0.80$. Para un intervalo de confianza del 98%, $\alpha = 0.02$. Al interpolar en la tabla A.6 encontramos $f_{0.01}(14,11) \approx 4.30$ y $f_{0.01}(11,14) \approx 3.87$. Por lo tanto, el intervalo de confianza del 98% para σ_1^2/σ_2^2 es

$$\left(\frac{3.07^2}{0.80^2}\right)\left(\frac{1}{4.30}\right) < \frac{\sigma_1^2}{\sigma_2^2} < \left(\frac{3.07^2}{0.80^2}\right)(3.87),$$

que se simplifica a $3.425 < \frac{\sigma_1^2}{\sigma_2^2} < 56.991$. Al calcular las raíces cuadradas de los límites de confianza encontramos que un intervalo de confianza del 98% para σ_1/σ_2 es

$$1.851 < \frac{\sigma_1}{\sigma_2} < 7.549.$$

Como este intervalo no permite la posibilidad de que σ_1/σ_2 sea igual a 1, es correcto suponer que $\sigma_1 \neq \sigma_2$ o $\sigma_1^2 \neq \sigma_2^2$ en el ejemplo 9.12.

Ejercicios

9.71 Un fabricante de baterías para automóvil afirma que sus baterías durarán, en promedio, 3 años con una varianza de 1 año. Suponga que 5 de estas baterías tienen duraciones de 1.9, 2.4, 3.0, 3.5 y 4.2 años y con base en esto construya un intervalo de confianza del 95% para σ^2, después decida si la afirmación del fabricante de que $\sigma^2 = 1$ es válida. Suponga que la población de duraciones de las baterías se distribuye de forma aproximadamente normal.

9.72 Una muestra aleatoria de 20 estudiantes obtuvo una media de $\bar{x} = 72$ y una varianza de $s^2 = 16$ en un examen universitario de colocación en matemáticas. Suponga que las calificaciones se distribuyen normalmente y con base en esto construya un intervalo de confianza del 98% para σ^2.

9.73 Construya un intervalo de confianza del 95% para σ^2 en el ejercicio 9.9 de la página 283.

9.74 Construya un intervalo de confianza del 99% para σ^2 en el ejercicio 9.11 de la página 283.

9.75 Construya un intervalo de confianza del 99% para σ en el ejercicio 9.12 de la página 283.

9.76 Construya un intervalo de confianza del 90% para σ en el ejercicio 9.13 de la página 283.

9.77 Construya un intervalo de confianza del 98% para σ_1/σ_2 en el ejercicio 9.42 de la página 295, donde σ_1 y σ_2 son, respectivamente, las desviaciones estándar para las distancias recorridas por litro de combustible de los camiones compactos Volkswagen y Toyota.

9.78 Construya un intervalo de confianza del 90% para σ_1^2/σ_2^2 en el ejercicio 9.43 de la página 295. ¿Se justifica que supongamos que $\sigma_1^2 \neq \sigma_2^2$ cuando construimos nuestro intervalo de confianza para $\mu_1 - \mu_2$?

9.79 Construya un intervalo de confianza del 90% para σ_1^2/σ_2^2 en el ejercicio 9.46 de la página 295. ¿Deberíamos suponer que $\sigma_1^2 = \sigma_2^2$ cuando construimos nuestro intervalo de confianza para $\mu_I - \mu_{II}$?

9.80 Construya un intervalo de confianza del 95% para σ_A^2/σ_B^2 en el ejercicio 9.49 de la página 295. ¿Tendría que utilizar la suposición de la igualdad de la varianza?

9.14 Estimación de la máxima verosimilitud (opcional)

A menudo los estimadores de parámetros han tenido que recurrir a la intuición. El estimador $\bar{X}$ ciertamente parece razonable como estimador de una media de la población μ. La virtud de S^2 como estimador de σ^2 se destaca en el estudio de estimadores insesgados de la sección 9.3. El estimador para un parámetro binomial p es simplemente una proporción de la muestra que, desde luego, es un *promedio* y recurre al sentido común. Sin embargo, hay muchas situaciones en las que no es del todo evidente cuál debería ser el estimador adecuado. Como resultado, el estudiante de estadística tiene mucho que aprender respecto a las diferentes filosofías que producen distintos métodos de estimación. En esta sección estudiaremos el **método de máxima verosimilitud**.

La estimación por máxima verosimilitud representa uno de los métodos de estimación más importantes en toda la estadística inferencial. No explicaremos el método de manera detallada; más bien, intentaremos transmitir la filosofía de la máxima verosimilitud e ilustrarla con ejemplos que la relacionan con otros problemas de estimación que se examinan en este capítulo.

Función de verosimilitud

Como el nombre lo indica, el método de máxima verosimilitud es aquel para el que se maximiza la *función de verosimilitud*, lo cual se ilustra mejor con un ejemplo que incluye una distribución discreta y un solo parámetro. Consideremos que $X_1, X_2,..., X_n$ son las variables aleatorias independientes tomadas de una distribución de probabilidad discreta representada por $f(\mathbf{x}, \theta)$, donde θ es un solo parámetro de la distribución. Ahora bien,

$$\begin{aligned} L(x_1, x_2, \ldots, x_n; \theta) &= f(x_1, x_2, \ldots, x_n; \theta) \\ &= f(x_1, \theta) f(x_2, \theta) \cdots f(x_n, \theta) \end{aligned}$$

es la *distribución conjunta de las variables aleatorias,* la cual a menudo se denomina función de probabilidad. Observe que la variable de la función de probabilidad es θ, no $\mathbf{x}$. Represente con $x_1, x_2,..., x_n$ los valores observados en una muestra. En el caso de una variable aleatoria discreta, la interpretación es muy clara. La cantidad $L(x_1, x_2,..., x_n; \theta)$, *la verosimilitud de la muestra*, es la siguiente probabilidad conjunta:

$$P(X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n \,|\, \theta),$$

que es la probabilidad de obtener los valores muestrales $x_1, x_2,..., x_n$. Para el caso discreto el estimador de máxima verosimilitud es el que da como resultado un valor máximo para esta probabilidad conjunta, o el que maximiza la probabilidad de la muestra.

Considere un ejemplo ficticio en el cual se inspeccionan tres artículos que salen de una línea de ensamble. Los artículos se clasifican como defectuosos o no defectuosos, de manera que se aplica el proceso de Bernoulli. La inspección de los tres artículos da como resultado dos artículos no defectuosos seguidos por uno defectuoso. Nos interesa estimar p, la proporción de artículos no defectuosos en el proceso. La probabilidad de la muestra para este ejemplo es dada por

$$p \cdot p \cdot q = p^2 q = p^2 - p^3,$$

donde $q = 1 - p$. La estimación de máxima verosimilitud daría un estimado de p para el que se maximiza la verosimilitud. Resulta claro que si diferenciamos la verosimilitud respecto a p, igualamos la derivada a cero y la resolvemos, obtenemos el valor

$$\hat{p} = \frac{2}{3}.$$

Entonces, desde luego, en esta situación $\hat{p} = 2/3$ es la proporción muestral defectuosa y, por ello, un estimador razonable de la probabilidad de un artículo defectuoso. El lector debería intentar comprender que la filosofía de la estimación de máxima verosimilitud proviene de la noción de que el estimador razonable de un parámetro que se basa en información muestral *es el valor del parámetro que produce la mayor probabilidad de obtener la muestra.* Ésta es, de hecho, la interpretación para el caso discreto, ya que la verosimilitud es la probabilidad de observar de manera conjunta los valores en la muestra.

Así, mientras que la interpretación de la función de verosimilitud como una probabilidad conjunta se limita al caso discreto, la noción de máxima verosimilitud se extiende a la estimación de parámetros de una distribución continua. Presentamos ahora una definición formal de la estimación de máxima verosimilitud.

Definición 9.3: Dadas las observaciones independientes $x_1, x_2, \ldots, x_n$ de una función de densidad de probabilidad (caso continuo) o de una función de masa de probabilidad (caso discreto) $f(\mathbf{x}, \theta)$, el estimador de máxima verosimilitud $\hat{\theta}$ es el que maximiza la función de probabilidad

$$L(x_1, x_2, \ldots, x_n; \theta) = f(\mathbf{x}; \theta) = f(x_1, \theta) f(x_2, \theta) \cdots f(x_n, \theta).$$

Muy a menudo conviene trabajar con el logaritmo natural de la función de verosimilitud para encontrar el máximo de esa función. Considere el siguiente ejemplo acerca del parámetro μ de una distribución de Poisson.

Ejemplo 9.20: Considere una distribución de Poisson con la siguiente función de masa de probabilidad

$$f(x|\mu) = \frac{e^{-\mu}\mu^x}{x!}, \quad x = 0, 1, 2, \ldots.$$

Suponga que se toma una muestra aleatoria $x_1, x_2, \ldots, x_n$ de la distribución. ¿Cuál es la estimación de máxima verosimilitud de μ?

***Solución*:** La función de probabilidad es

$$L(x_1, x_2, \ldots, x_n; \mu) = \prod_{i=1}^{n} f(x_i|\mu) = \frac{e^{-n\mu}\mu^{\sum_{i=1}^{n} x_i}}{\prod_{i=1}^{n} x_i!}.$$

Considere ahora

$$\ln L(x_1, x_2, \ldots, x_n; \mu) = -n\mu + \sum_{i=1}^{n} x_i \ln \mu - \ln \prod_{i=1}^{n} x_i!$$

$$\frac{\partial \ln L(x_1, x_2, \ldots, x_n; \mu)}{\partial \mu} = -n + \sum_{i=1}^{n} \frac{x_i}{\mu}.$$

Resolver para $\hat{\mu}$, el estimador de máxima verosimilitud, implica definir la derivada para cero y resolver para el parámetro. Por consiguiente,

$$\hat{\mu} = \sum_{i=1}^{n} \frac{x_i}{n} = \bar{x}.$$

La segunda derivada de la función de verosimilitud logarítmica es negativa, lo cual implica que la solución anterior realmente es un máximo. Como μ es la media de la distribución de Poisson (capítulo 5), el promedio muestral en realidad parecería ser un estimador razonable. ▟

El siguiente ejemplo presenta el uso del método de máxima verosimilitud para calcular estimados de dos parámetros. Simplemente encontramos los valores de los parámetros que maximizan (de forma conjunta) la función de probabilidad.

Ejemplo 9.21: Considere una muestra aleatoria $x_1, x_2, \ldots, x_n$ de una distribución normal $N(\mu, \sigma)$. Calcule los estimadores de máxima verosimilitud para μ y σ^2.

***Solución*:** La función de verosimilitud para la distribución normal es

$$L(x_1, x_2, \ldots, x_n; \mu, \sigma^2) = \frac{1}{(2\pi)^{n/2}(\sigma^2)^{n/2}} \exp\left[-\frac{1}{2}\sum_{i=1}^{n}\left(\frac{x_i - \mu}{\sigma}\right)^2\right].$$

Al usar logaritmos obtenemos

$$\ln L(x_1, x_2, \ldots, x_n; \mu, \sigma^2) = -\frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln\sigma^2 - \frac{1}{2}\sum_{i=1}^{n}\left(\frac{x_i - \mu}{\sigma}\right)^2.$$

Por lo tanto,

$$\frac{\partial \ln L}{\partial \mu} = \sum_{i=1}^{n}\left(\frac{x_i - \mu}{\sigma^2}\right)$$

y

$$\frac{\partial \ln L}{\partial \sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{2(\sigma^2)^2}\sum_{i=1}^{n}(x_i - \mu)^2.$$

Al igualar ambas derivadas a cero, obtenemos

$$\sum_{i=1}^{n} x_i - n\mu = 0 \quad \text{y} \quad n\sigma^2 = \sum_{i=1}^{n}(x_i - \mu)^2.$$

Por consiguiente, el estimador de máxima verosimilitud de μ es dado por

$$\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} x_i = \bar{x},$$

que es un resultado satisfactorio, ya que $\bar{x}$ ha desempeñado un papel tan importante en este capítulo como un estimador puntual de μ. Por otro lado, el estimador de máxima verosimilitud de σ^2 es

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2.$$

Al verificar la matriz derivada parcial de segundo orden se confirma que la solución da como resultado el máximo de la función de verosimilitud. ▟

Resulta interesante notar la distinción entre el estimador de máxima verosimilitud de σ^2 y el estimador insesgado S^2 que se presentó al principio de este capítulo. Los numeradores son idénticos, desde luego, y el denominador lo constituyen los "grados de libertad" $n - 1$ para el estimador insesgado, y n para el estimador de máxima verosimilitud. Los estimadores de máxima verosimilitud no necesariamente gozan de la propiedad de carecer de sesgo. Sin embargo, los estimadores de máxima verosimilitud tienen importantes propiedades asintóticas.

Ejemplo 9.22: Suponga que en un estudio biomédico se utilizan 10 ratas a las que después de inyectarles células cancerosas se les suministra un fármaco contra el cáncer diseñado para aumentar su tasa de supervivencia. Los tiempos de supervivencia, en meses, son 14, 17, 27, 18, 12, 8, 22, 13, 19 y 12. Suponga que se trata de una distribución exponencial. Calcule un estimado de máxima verosimilitud de la supervivencia media.

***Solución*:** Del capítulo 6 sabemos que la función de densidad de probabilidad para la variable aleatoria exponencial X es

$$f(x, \beta) = \begin{cases} \frac{1}{\beta} e^{-x/\beta}, & x > 0, \\ 0, & \text{en cualquier caso.} \end{cases}$$

Por consiguiente, la función de verosimilitud logarítmica de los datos, dado que $n = 10$, es

$$\ln L(x_1, x_2, \ldots, x_{10}; \beta) = -10 \ln \beta - \frac{1}{\beta} \sum_{i=1}^{10} x_i.$$

Si se establece que

$$\frac{\partial \ln L}{\partial \beta} = -\frac{10}{\beta} + \frac{1}{\beta^2} \sum_{i=1}^{10} x_i = 0$$

implica que

$$\hat{\beta} = \frac{1}{10} \sum_{i=1}^{10} x_i = \bar{x} = 16.2.$$

Si se evalúa la segunda derivada de la función de verosimilitud logarítmica en el valor $\hat{\beta}$ anterior se produce un valor negativo. Como resultado, el estimador del parámetro β, la media de la población, es el promedio muestral $\bar{x}$.

El siguiente ejemplo ilustra el estimador de máxima verosimilitud para una distribución que no se incluye en los capítulos anteriores.

Ejemplo 9.23: Se sabe que una muestra que consta de los valores 12, 11.2, 13.5, 12.3, 13.8 y 11.9 proviene de una población con la siguiente función de densidad

$$f(x; \theta) = \begin{cases} \frac{\theta}{x^{\theta+1}}, & x > 1, \\ 0, & \text{en cualquier caso,} \end{cases}$$

donde $\theta > 0$. Calcule la estimación de máxima verosimilitud de θ.

***Solución*:** La función de verosimilitud de n observaciones de esta población se escribe como

$$L(x_1, x_2, \ldots, x_{10}; \theta) = \prod_{i=1}^{n} \frac{\theta}{x_i^{\theta+1}} = \frac{\theta^n}{(\prod_{i=1}^{n} x_i)^{\theta+1}},$$

lo cual implica que

$$\ln L(x_1, x_2, \ldots, x_{10}; \theta) = n \ln(\theta) - (\theta + 1) \sum_{i=1}^{n} \ln(x_i).$$

Si establecemos que $0 = \frac{\partial \ln L}{\partial \theta} = \frac{n}{\theta} - \sum_{i=1}^{n} \ln(x_i)$ da como resultado

$$\hat{\theta} = \frac{n}{\sum_{i=1}^{n} \ln(x_i)}$$

$$= \frac{6}{\ln(12) + \ln(11.2) + \ln(13.5) + \ln(12.3) + \ln(13.8) + \ln(11.9)} = 0.3970.$$

Como la segunda derivada de L es $-n/\theta^2$, que siempre es negativa, la función de probabilidad alcanza su valor máximo en $\hat{\theta}$.

Comentarios adicionales respecto a la estimación de máxima verosimilitud

Un análisis detallado de las propiedades de la estimación de máxima verosimilitud está fuera del alcance de este libro y, por lo general, es un tema importante en un curso teórico de estadística inferencial. El método de máxima verosimilitud permite al analista utilizar el conocimiento de la distribución para determinar un estimador adecuado. *El método de máxima verosimilitud no se puede aplicar si no se conoce la distribución subyacente.* En el ejemplo 9.21 aprendimos que el estimador de máxima verosimilitud no necesariamente carece de sesgo. El estimador de máxima verosimilitud es insesgado *asintóticamente* o *en el límite*; es decir, la magnitud del sesgo se aproxima a cero a medida que la muestra se hace más grande. Al principio de este capítulo examinamos la noción de eficacia, que se vincula con la propiedad de la varianza de un estimador. Los estimadores de máxima verosimilitud tienen propiedades de varianza deseables en el límite. El lector debería consultar la obra de Lehmann y D'Abrera (1998) para más detalles.

Ejercicios

9.81 Suponga que hay n ensayos $x_1, x_2, \ldots, x_n$ de un proceso de Bernoulli con parámetro p, la probabilidad de un éxito. Esto es, la probabilidad de r éxitos es dada por $\binom{n}{r} p^r (1-p)^{n-r}$. Determine el estimador de máxima verosimilitud para el parámetro p.

9.82 Considere la distribución logarítmica normal con la función de densidad dada en la sección 6.9. Suponga que tiene una muestra aleatoria $x_1, x_2, \ldots, x_n$ de una distribución logarítmica normal.

a) Escriba la función de verosimilitud.

b) Desarrolle los estimadores de máxima verosimilitud de μ y σ^2.

9.83 Considere una muestra aleatoria de $x_1, \ldots, x_n$ obtenida de la distribución gamma descrita en la sección 6.6. Suponga que conoce el parámetro α, el cual digamos que es 5, y con base en esto determine la estimación de máxima verosimilitud para el parámetro β.

9.84 Considere una muestra aleatoria de $x_1, x_2, \ldots, x_n$ observaciones de una distribución de Weibull con parámetros α y β, y la siguiente función de densidad

$$f(x) = \begin{cases} \alpha\beta x^{\beta-1} e^{-\alpha x^{\beta}}, & x > 0, \\ 0, & \text{en cualquier caso,} \end{cases}$$

para $\alpha, \beta > 0$.

a) Escriba la función de verosimilitud.

b) Escriba las ecuaciones que al resolverse proporcionan los estimadores de máxima verosimilitud de α y β.

9.85 Considere una muestra aleatoria de $x_1, \ldots, x_n$ obtenida de una distribución uniforme $U(0, \theta)$, con el parámetro θ desconocido, donde $\theta > 0$. Determine el estimador de máxima verosimilitud de θ.

9.86 Considere las observaciones independientes de $x_1, x_2, \ldots, x_n$ de la distribución gamma que se analizó en la sección 6.6.

a) Escriba la función de verosimilitud.
b) Escriba un conjunto de ecuaciones que, cuando se resuelven, proporcionan los estimadores de máxima verosimilitud de α y β.

9.87 Considere un experimento hipotético en el que un hombre que tiene un hongo utiliza un medicamento fungicida y se cura. Por lo tanto, considere que se trata de una muestra de una distribución de Bernoulli con la siguiente función de probabilidad

$$f(x) = p^x q^{1-x}, \quad x = 0, 1,$$

donde p es la probabilidad de un éxito (curación) y $q = 1 - p$. Ahora, desde luego, la información muestral da $x = 1$. Escriba un procedimiento que demuestre que $\hat{p} = 1.0$ es el estimador de máxima probabilidad de curación.

9.88 Considere la observación X de la distribución binomial negativa dada en la sección 5.4. Calcule el estimador de máxima verosimilitud para p, suponiendo que se conoce k.

Ejercicios de repaso

9.89 Considere dos estimadores de σ^2 para una muestra $x_1, x_2, \ldots, x_n$ que se extrae de una distribución normal con media μ y varianza σ^2. Los estimadores son el estimador insesgado $s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$ y el estimador de máxima verosimilitud $\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2$. Analice las propiedades de la varianza de estos dos estimadores.

9.90 De acuerdo con el *Roanoke Times*, McDonald's vendió 42.1% de la participación del mercado de hamburguesas. Una muestra aleatoria de 75 hamburguesas vendidas reveló que 28 de ellas fueron vendidas por McDonald's. Utilice el material de la sección 9.10 para determinar si esta información respalda la afirmación del *Roanoke Times*.

9.91 Se afirma que un individuo podrá reducir, en un lapso de 2 semanas, un promedio de 4.5 kilogramos de peso con una nueva dieta. Los pesos de 7 mujeres que siguieron esta dieta se registraron antes y después de un periodo de 2 semanas.

Mujer	Peso antes	Peso después
1	58.5	60.0
2	60.3	54.9
3	61.7	58.1
4	69.0	62.1
5	64.0	58.5
6	62.6	59.9
7	56.7	54.4

Pruebe la afirmación sobre la dieta calculando un intervalo de confianza del 95% para la diferencia media en el peso. Suponga que las diferencias de los pesos se distribuyen de forma aproximadamente normal.

9.92 En Virginia Tech se realizó un estudio para determinar si se puede utilizar el fuego como una herramienta de control viable para aumentar la cantidad de forraje disponible para los venados durante los meses críticos a finales del invierno y principios de la primavera. El calcio es un elemento necesario para las plantas y los animales. La cantidad que la planta toma y almacena está estrechamente correlacionada con la cantidad presente en el suelo. Se formuló la hipótesis de que el fuego podría cambiar los niveles de calcio presentes en el suelo y, por lo tanto, influir en la cantidad disponible para los venados. Se seleccionó una extensión grande de tierra en el bosque Fishburn para provocar un incendio controlado. Justo antes de la quema se tomaron muestras de suelo de 12 parcelas con la misma área y se analizaron para verificar su contenido de calcio. Después del incendio se volvieron a analizar los niveles de calcio en las mismas parcelas. Los valores obtenidos, en kilogramos por parcela, se presentan en la siguiente tabla:

Nivel de calcio (kg/parcela)		
Parcela	**Antes del incendio**	**Después del incendio**
1	50	9
2	50	18
3	82	45
4	64	18
5	82	18
6	73	9
7	77	32
8	54	9
9	23	18
10	45	9
11	36	9
12	54	9

Construya un intervalo de confianza del 95% para la diferencia media en los niveles de calcio presentes en el suelo antes y después del incendio controlado. Suponga que la distribución de las diferencias en los niveles de calcio es aproximadamente normal.

9.93 El dueño de un gimnasio afirma que una persona podrá reducir, en un periodo de 5 días, un promedio de 2 centímetros en su talla de cintura con un nuevo programa de ejercicios. En la siguiente tabla se presentan las tallas de cintura de 6 hombres que participaron en este programa de ejercicios antes y después del periodo de 5 días:

Hombre	Talla de cintura antes	Talla de cintura después
1	90.4	91.7
2	95.5	93.9
3	98.7	97.4
4	115.9	112.8
5	104.0	101.3
6	85.6	84.0

Mediante el cálculo de un intervalo de confianza del 95% para la reducción media en la talla de cintura determine si la afirmación del dueño del gimnasio es válida. Suponga que la distribución de las diferencias en las tallas de cintura antes y después del programa es aproximadamente normal.

9.94 El Departamento de Ingeniería Civil del Virginia Tech comparó una técnica de ensayo modificada (M-5 hr) para recuperar coliformes fecales en residuos líquidos (charcos) de agua de lluvia en una área urbana con la técnica del número más probable (NMP). El departamento recolectó un total de 12 muestras de tales residuos y las analizó con las dos técnicas. Los conteos de coliformes fecales por 100 mililitros se registraron en la siguiente tabla:

Muestra	Conteo NMP	Conteo con M-5 hr
1	2300	2010
2	1200	930
3	450	400
4	210	436
5	270	4100
6	450	2090
7	154	219
8	179	169
9	192	194
10	230	174
11	340	274
12	194	183

Construya un intervalo de confianza del 90% para la diferencia entre el conteo medio de coliformes fecales que se obtuvo con la técnica M-5 hr y el que se obtuvo con la NMP. Suponga que las diferencias en los conteos se distribuyen de forma aproximadamente normal.

9.95 Se llevó a cabo un experimento para determinar si el acabado superficial tiene un efecto en el límite de resistencia a la fatiga del acero. Una teoría indica que el pulido aumenta el límite medio de resistencia a la fatiga (para la flexión inversa). Desde un punto de vista práctico, el pulido no debería tener efecto alguno sobre la desviación estándar del límite de resistencia a la fatiga, el cual se sabe, a partir de la realización de diversos experimentos de límite de resistencia a la fatiga, que es de 4000 psi. Se realiza un experimento sobre acero al carbono al 0.4% usando especímenes sin pulido y especímenes con pulido suave. Los datos son los siguientes:

Límite de fatiga (psi)	
Acero al carbono al 0.4%	**Acero al carbono al 0.4% sin pulir**
85,500	82,600
91,900	82,400
89,400	81,700
84,000	79,500
89,900	79,400
78,700	69,800
87,500	79,900
83,100	83,400

Calcule un intervalo de confianza del 95% para la diferencia entre las medias de la población para los dos métodos. Suponga que las poblaciones se distribuyen de forma aproximadamente normal.

9.96 Un antropólogo está interesado en determinar la proporción de individuos de dos tribus indias que tienen doble remolino de cabello en la zona occipital. Suponga que toma muestras independientes de cada una de las dos tribus y encuentra que 24 de 100 individuos de la tribu A y 36 de 120 individuos de la tribu B poseen tal característica. Construya un intervalo de confianza del 95% para la diferencia $p_B - p_A$ entre las proporciones de estas dos tribus con remolinos de cabello en la zona occipital.

9.97 Un fabricante de planchas eléctricas produce estos artículos en dos plantas en las que las partes pequeñas son surtidas por el mismo proveedor. El fabricante puede ahorrar algo si le compra a un proveedor local los termostatos para la planta B. Para probar si estos nuevos termostatos son tan precisos como los anteriores le compra sólo un lote al proveedor local y los prueba en planchas a 550°F. Al final lee con un termopar las temperaturas reales y las redondea al siguiente 0.1°F más cercano. Los datos son los siguientes:

Proveedor nuevo (°F)					
530.3	559.3	549.4	544.0	551.7	566.3
549.9	556.9	536.7	558.8	538.8	543.3
559.1	555.0	538.6	551.1	565.4	554.9
550.0	554.9	554.7	536.1	569.1	

Proveedor anterior (°F)					
559.7	534.7	554.8	545.0	544.6	538.0
550.7	563.1	551.1	553.8	538.8	564.6
554.5	553.0	538.4	548.3	552.9	535.1
555.0	544.8	558.4	548.7	560.3	

Calcule un intervalo de confianza de 95% para σ_1^2/σ_2^2 y para σ_1/σ_2, donde σ_1^2 y σ_2^2 son las varianzas de la población de las lecturas de los termostatos del proveedor nuevo y del anterior, respectivamente.

9.98 Se afirma que la resistencia del alambre *A* es mayor que la del alambre *B*. Un experimento sobre los alambres muestra los siguientes resultados (en ohms):

Alambre *A*	**Alambre *B***
0.140	0.135
0.138	0.140
0.143	0.136
0.142	0.142
0.144	0.138
0.137	0.140

Suponga varianzas iguales y explique a qué conclusiones llega si se basa en esto.

9.99 Una forma alternativa de estimación se lleva a cabo a través del método de momentos. El método consiste en igualar la media y la varianza de la población con las correspondientes media muestral $\bar{x}$ y varianza muestral s^2, y resolver para los parámetros; el resultado son los **estimadores por momentos**. En el caso de un solo parámetro sólo se utilizan las medias. Argumente por qué en el caso de la distribución de Poisson el estimador de máxima verosimilitud y los estimadores por momentos son iguales.

9.100 Especifique los estimadores por momentos para μ y σ^2 para la distribución normal.

9.101 Especifique los estimadores por momentos para μ y σ^2 para la distribución logarítmica normal.

9.102 Especifique los estimadores por momentos para α y β en el caso de la distribución gamma.

9.103 Se realizó una encuesta con el fin de comparar los sueldos de administradores de plantas químicas empleados en dos áreas del país: el norte y el centro-occidente. Se eligió una muestra aleatoria independiente de 300 gerentes de planta para cada una de las dos áreas. A tales gerentes se les preguntó el monto de su sueldo anual. Los resultados fueron los siguientes:

Norte	**Centro-Occidente**
$\bar{x}_1 = \$102,300$	$\bar{x}_2 = \$98,500$
$s_1 = \$5700$	$s_2 = \$3800$

a) Construya un intervalo de confianza del 99% para $\mu_1 - \mu_2$, la diferencia en los sueldos medios.

b) ¿Qué supuso en el inciso *a*) acerca de la distribución de los sueldos anuales para las dos áreas? ¿Es necesaria la suposición de normalidad? Explique su respuesta.

c) ¿Qué supuso acerca de las dos varianzas? ¿Es razonable la suposición de igualdad de varianzas? ¡Explique!

9.104 Considere el ejercicio de repaso 9.103. Suponga que los datos aún no se han recabado. Suponga también que los estadísticos previos sugieren que $\sigma_1 = \sigma_2 = \$4000$. ¿Los tamaños de las muestras en el ejercicio de repaso 9.103 son suficientes para producir un intervalo de confianza del 95% si $\mu_1 - \mu_2$ tiene una anchura de sólo \$1000? Presente el desarrollo completo.

9.105 Un sindicato se preocupa por el notorio ausentismo de sus miembros. Los líderes del sindicato siempre habían afirmado que, en un mes típico, el 95% de sus afiliados estaban ausentes menos de 10 horas al mes. El sindicato decide verificar esto revisando una muestra aleatoria de 300 de sus miembros. Se registra el número de horas de ausencia para cada uno de los 300 miembros. Los resultados son $\bar{x} = 6.5$ horas y $s = 2.5$ horas. Utilice los datos para responder esa afirmación utilizando un límite de tolerancia unilateral y eligiendo un nivel de confianza del 99%. Asegúrese de aplicar lo que ya sabe acerca del cálculo del límite de tolerancia.

9.106 Se seleccionó una muestra aleatoria de 30 empresas que comercializan productos inalámbricos para determinar la proporción de tales empresas que implementaron software nuevo para aumentar la productividad. Resultó que 8 de las 30 empresas habían implementado tal software. Calcule un intervalo de confianza del 95% en *p*, la proporción verdadera de ese tipo de empresas que implementaron el nuevo software.

9.107 Remítase al ejercicio de repaso 9.106. Suponga que se desea saber si la estimación puntual $\hat{p} = 8/30$ es lo suficientemente precisa porque el intervalo de confianza alrededor de *p* no es tan estrecho como se requiere. Utilice $\hat{p}$ como el estimado de *p* para determinar cuántas empresas habría que incluir en una muestra para obtener un intervalo de confianza del 95% con una anchura de sólo 0.05.

9.108 Un fabricante produce un artículo que se clasifica como "defectuoso" o "no defectuoso". Para estimar la proporción de productos defectuosos se tomó una muestra aleatoria de 100 artículos de la producción y se encontraron 10 defectuosos. Después de aplicar un programa de mejoramiento de la calidad se volvió a realizar el experimento. Se tomó una nueva muestra de 100 artículos y esta vez sólo 6 salieron defectuosos.

a) Dado un intervalo de confianza del 95% de $p_1 - p_2$, donde p_1 y p_2 representan la proporción de artículos defectuosos de la población antes y después del mejoramiento, respectivamente.

b) ¿Hay información en el intervalo de confianza que se encontró en el inciso *a*) que sugiera que $p_1 > p_2$? Explique su respuesta.

9.109 Se utiliza una máquina para llenar cajas de un producto en una operación de la línea de ensamble. Gran parte del interés se centra en la variabilidad del número de onzas del producto en la caja. Se sabe que la desviación estándar en el peso del producto es de 0.3 onzas. Se realizan mejoras y luego se toma una muestra aleatoria de 20 cajas, y se encuentra que la varianza de la muestra es de 0.045 onzas^2. Calcule un intervalo de confianza del 95% de la varianza del peso del producto. Si considera el rango del intervalo de confianza, ¿le parece que el mejoramiento en el proceso incrementó la calidad en lo que se refiere a la variabilidad? Suponga normalidad en la distribución del peso del producto.

9.110 Un grupo de consumidores está interesado en comparar los costos de operación de dos diferentes tipos de motor para automóvil. El grupo encuentra 15 propietarios cuyos automóviles tienen motor tipo *A* y 15 que tienen motor tipo *B*. Los 30 propietarios compraron sus automóviles más o menos al mismo tiempo y todos llevaron buenos registros en cierto periodo de 12 meses. Los consumidores encontraron, además, que los propietarios recorrieron aproximadamente el mismo número de millas. Los estadísticos de costo son $\bar{y}_A = \$87.00/1000$ millas, $\bar{y}_B = \$75.00/1000$ millas, $s_A = \$5.99$ y $s_B = \$4.85$. Calcule un intervalo de confianza del 95% para estimar $\mu_A - \mu_B$, la diferencia en el costo medio de operación. Suponga normalidad y varianzas iguales.

9.111 Considere el estadístico S_p^2, el estimado agrupado de σ^2 que se estudió en la sección 9.8 y que se utiliza cuando se está dispuesto a suponer que $\sigma_1^2 = \sigma_2^2 = \sigma^2$. Demuestre que el estimador es insesgado para σ^2 [es decir, demuestre que $E(S_p^2) = \sigma^2$]. Puede utilizar los resultados de cualquier teorema o ejemplo de este capítulo.

9.112 Un grupo de investigadores del factor humano están interesados en saber cómo reaccionan los pilotos aviadores ante un estímulo dispuesto de cierta manera en la cabina del avión. Para lograr su objetivo realizaron un experimento de simulación en un laboratorio, el cual incluyó a 15 pilotos, los que presentaron un tiempo de reacción promedio de 3.2 segundos y una desviación estándar muestral de 0.6 segundos. Resulta de interés caracterizar el extremo, es decir, el escenario del peor caso. Para conseguir esto realice lo siguiente:

a) Determine un importante límite de confianza unilateral específico del 99% del tiempo medio de reacción. ¿Qué suposición, si la hubiera, debería hacer acerca de la distribución de los tiempos de reacción?

b) Determine un intervalo unilateral de predicción del 99% e interprete su significado. ¿Debería usted suponer algo sobre la distribución de los tiempos de reacción para calcular este límite?

c) Calcule un límite de tolerancia unilateral con una confianza del 99% que incluya al 95% de los tiempos de reacción. Nuevamente, de ser necesario, interprete o suponga algo acerca de la distribución. [Nota: Los valores del límite de tolerancia unilateral también se incluyen en la tabla A.7].

9.113 Cierto proveedor fabrica un tipo de tapete de hule que vende a las empresas automotrices. El material que utiliza para los tapetes debe tener ciertas características de dureza. Ocasionalmente detecta tapetes defectuosos en el proceso y los rechaza. El proveedor afirma que la proporción de tapetes defectuosos es de 0.05, pero como un cliente que compró los tapetes desafió su afirmación, realizó un experimento en el que se probaron 400 tapetes y se encontraron 17 defectuosos.

a) Calcule un intervalo de confianza bilateral del 95% de la proporción de tapetes defectuosos.

b) Calcule un intervalo de confianza unilateral del 95% adecuado de la proporción de tapetes defectuosos.

c) Interprete los intervalos de ambos incisos y comente acerca de la afirmación hecha por el proveedor.

9.15 Posibles riesgos y errores conceptuales: relación con el material de otros capítulos

El concepto de *intervalo de confianza de muestra grande* en una población a menudo confunde a los alumnos principiantes. Se basa en la idea de que incluso cuando se desconoce σ y no se está convencido de que la distribución que se muestrea es normal, se puede calcular un intervalo de confianza para μ a partir de

$$\bar{x} \pm z_{\alpha/2} \frac{s}{\sqrt{n}}.$$

En la práctica es común que se utilice esta fórmula cuando la muestra es demasiado pequeña. El origen de este intervalo de muestra grande es, por supuesto, el teorema del límite central (TLC), con el cual la normalidad no es necesaria. Aquí el TLC requiere una σ conocida, de la cual s sólo es un estimado. Por lo tanto, n debe ser al menos tan grande como 30 y la distribución subyacente debe tener una simetría similar, en cuyo caso el intervalo sigue siendo una aproximación.

Hay casos en que la aplicación práctica del material de este capítulo depende en gran medida del contexto específico. Un ejemplo muy importante es el uso de la distribución t para el intervalo de confianza de μ cuando se desconoce σ. En términos estrictos, el uso de la distribución t requiere que la distribución de donde se toma la muestra sea normal. Sin embargo, es bien sabido que cualquier aplicación de la distribución t es razonablemente insensible, es decir, **robusta**, a la suposición de normalidad. Esto representa una de esas situaciones afortunadas que ocurren con frecuencia en el campo de la estadística, donde no se sostiene un supuesto básico y "¡todo resulta bien!" Sin embargo, la población de la que se toma la muestra no se puede desviar mucho de la normalidad. Por consiguiente, a menudo se recurrirá a las gráficas de probabilidad normal estudiadas en el capítulo 8 y las pruebas de bondad del ajuste que se presentarán en el capítulo 10 para atribuir algún sentido de "cercanía a la normalidad". Esta idea de "robustez a la normalidad" se volverá a presentar en el capítulo 10.

Por experiencia sabemos que uno de los más graves "usos incorrectos de la estadística" en la práctica surge de la confusión sobre las diferencias en la interpretación de los tipos de intervalos estadísticos. Por consiguiente, la subsección de este capítulo en la que se examinan las diferencias entre los tres tipos de intervalos es importante. Es muy probable que en la práctica **se utilice en exceso el intervalo de confianza**, es decir, que se emplee cuando no es la media lo que interesa en realidad, sino la cuestión de: "¿en dónde va a caer la siguiente observación?", o la a menudo más importante cuestión de: "¿en dónde se ubica la mayor parte de la distribución?" Éstas son preguntas fundamentales que no se pueden responder calculando un intervalo de la media. A menudo resulta confusa la interpretación de un intervalo de confianza. Es tentador concluir que hay una probabilidad de 0.95 de que el parámetro caiga dentro del intervalo. Aunque se trata de una interpretación correcta del **intervalo posterior bayesiano** (para mayores referencias sobre la inferencia bayesiana véase el capítulo 18), no es una interpretación adecuada de la frecuencia.

El intervalo de confianza tan sólo sugiere que si se realiza el experimento y los datos se observan una y otra vez, aproximadamente 95% de tales intervalos contendrá el parámetro verdadero. Cualquier alumno principiante de la estadística práctica debería tener muy claras las diferencias entre estos intervalos estadísticos.

Otro posible y grave uso incorrecto de la estadística es el que se cometería si se aplicara la distribución χ^2 a un intervalo de confianza de una sola varianza. De nuevo, se supone normalidad en la distribución de donde se toma la muestra. A diferencia del resultado de utilizar la distribución t, la prueba χ^2 para esta aplicación **no es robusta para la suposición de normalidad** (esto significa que cuando la distribución subyacente no es normal, la distribución muestral de $\frac{(n-1)S^2}{\sigma^2}$ se aparta mucho de χ^2). En consecuencia, el uso estricto de la prueba de bondad de ajuste (véase el capítulo 10) y de las gráficas de probabilidad normal, o de la prueba y las gráficas, puede ser muy importante en esos contextos. En los siguientes capítulos se proporcionará más información sobre este tema general.

Capítulo 10

Pruebas de hipótesis de una y dos muestras

10.1 Hipótesis estadísticas: conceptos generales

Como se expuso en el capítulo 9, a menudo el problema al que se enfrentan el científico o el ingeniero no es tanto la estimación de un parámetro de la población, sino la formación de un procedimiento de decisión que se base en los datos y que pueda producir una conclusión acerca de algún sistema científico. Por ejemplo, un investigador médico puede decidir con base en evidencia experimental si beber café incrementa el riesgo de cáncer en los seres humanos; un ingeniero quizá tenga que decidir con base en datos muestrales si hay una diferencia entre la precisión de un tipo de medidor y la de otro; o tal vez un sociólogo desee reunir los datos apropiados que le permitan decidir si el tipo de sangre y el color de ojos de un individuo son variables independientes. En cada uno de estos casos el científico o el ingeniero *postulan* o *conjeturan* algo acerca de un sistema. Además, cada uno debe utilizar datos experimentales y tomar decisiones basadas en ellos. En cada caso la conjetura se puede expresar en forma de hipótesis estadística. Los procedimientos que conducen a la aceptación o al rechazo de hipótesis estadísticas como éstas comprenden una área importante de la inferencia estadística. Empecemos por definir con precisión lo que entendemos por **hipótesis estadística**.

Definición 10.1: Una **hipótesis estadística** es una aseveración o conjetura respecto a una o más poblaciones.

La verdad o falsedad de una hipótesis estadística nunca se sabe con absoluta certeza, a menos que se examine toda la población, lo cual, por supuesto, sería poco práctico en la mayoría de las situaciones. En vez de eso se toma una muestra aleatoria de la población de interés y se utilizan los datos contenidos en ella para proporcionar evidencia que respalde o no la hipótesis. La evidencia de la muestra que es inconsistente con la hipótesis planteada conduce al rechazo de la misma.

El papel que desempeña la probabilidad en la prueba de hipótesis

Debería quedar claro al lector que un procedimiento de toma de decisiones debe implicar la conciencia de la *probabilidad de llegar a una conclusión errónea*. Por ejemplo, suponga que la hipótesis que postuló el ingeniero es que la fracción p de artículos defectuosos en cierto proceso es 0.10. El experimento consiste en observar una muestra aleatoria del producto en cuestión. Suponga que se prueban 100 artículos y que se encuentran 12 defectuosos. Es razonable concluir que esta evidencia no rechaza la condición de que el parámetro binomial $p = 0.10$, por lo que puede provocar que no se rechace la hipótesis. Sin embargo, también puede provocar que no se refute $p = 0.12$, o quizá incluso $p = 0.15$. Como resultado, el lector se debe acostumbrar a la idea de que **el rechazo de una hipótesis implica que fue refutada por la evidencia de la muestra**. En otras palabras, **el rechazo significa que existe una pequeña probabilidad de obtener la información muestral observada cuando, de hecho, la hipótesis es verdadera**. Por ejemplo, en la hipótesis de la proporción de artículos defectuosos, una muestra de 100 artículos que revela que hay 20 defectuosos es ciertamente evidencia para el rechazo. ¿Por qué? Si en realidad $p = 0.10$, la probabilidad de obtener 20 o más artículos defectuosos es aproximadamente de 0.002. Con el pequeño riesgo resultante de llegar a una conclusión errónea parecería seguro **rechazar la hipótesis** de que $p = 0.10$. En otras palabras, el rechazo de una hipótesis tiende a casi "descartar" la hipótesis. Por otro lado, es muy importante enfatizar que la aceptación o, más bien, la falta de rechazo no descarta otras posibilidades. Como resultado, *el analista de datos establece una conclusión firme cuando se rechaza una hipótesis*.

En el planteamiento formal de una hipótesis a menudo influye la estructura de la probabilidad de una conclusión errónea. Si el científico está interesado en *apoyar firmemente* un argumento, espera llegar a éste en la forma del rechazo de una hipótesis. Si el investigador médico desea mostrar evidencia sólida a favor del argumento de que beber café aumenta el riesgo de contraer cáncer, la hipótesis a probar debería tener la forma "el riesgo de desarrollar cáncer no aumenta como consecuencia de beber café". Como resultado, el argumento se obtiene mediante un rechazo. De manera similar, para apoyar la afirmación de que un tipo de medidores es más preciso que otro, el ingeniero prueba la hipótesis de que no hay diferencia en la precisión de los dos tipos de medidores.

Lo anterior implica que cuando el analista de datos formaliza la evidencia experimental con base en la prueba de hipótesis, es muy importante el **planteamiento formal de la hipótesis**.

La hipótesis nula y la hipótesis alternativa

La estructura de la prueba de hipótesis se establece usando el término **hipótesis nula**, el cual se refiere a cualquier hipótesis que se desea probar y se denota con H_0. El rechazo de H_0 conduce a la aceptación de una **hipótesis alternativa**, que se denota con H_1. La comprensión de las diferentes funciones que desempeñan la hipótesis nula (H_0) y la hipótesis alternativa (H_1) es fundamental para entender los principios de la prueba de hipótesis. La hipótesis alternativa H_1 por lo general representa la *pregunta que se responderá o la teoría que se probará*, por lo que su especificación es muy importante. La hipótesis nula H_0 *anula o se opone a* H_1 y a menudo es el complemento lógico de H_1. A medida que el lector aprenda más sobre la prueba de hipótesis notará que el analista llega a una de las siguientes dos conclusiones:

rechazar H_0 a favor de H_1 debido a evidencia suficiente en los datos o
no rechazar H_0 debido a evidencia insuficiente en los datos.

Observe que las *conclusiones no implican una "aceptación de H_0" formal y literal.* La aseveración de H_0 a menudo representa el "status quo" contrario a una nueva idea, conjetura, etcétera, enunciada en H_1; en tanto que no rechazar H_0 representa la conclusión adecuada. En nuestro ejemplo binomial la cuestión práctica podría ser el interés en que la probabilidad histórica de artículos defectuosos de 0.10 ya no sea verdadera. De hecho, la conjetura podría ser que p excede a 0.10. Entonces podríamos afirmar que

$$H_0\colon p = 0.10,$$
$$H_1\colon p > 0.10.$$

Ahora, 12 artículos defectuosos de cada 100 no refutan $p = 0.10$, por lo que la conclusión es "no rechazar H_0". Sin embargo, si los datos revelan 20 artículos defectuosos de cada 100, la conclusión sería "rechazar H_0" a favor de H_1: $p > 0.10$.

Aunque las aplicaciones de la prueba de hipótesis son muy abundantes en trabajos científicos y de ingeniería, quizás el mejor ejemplo para un principiante sea el dilema que enfrenta el jurado en un juicio. Las hipótesis nula y alternativa son

H_0: el acusado es inocente,
H_1: el acusado es culpable.

La acusación proviene de una sospecha de culpabilidad. La hipótesis H_0 (el status quo) se establece en oposición a H_1 y se mantiene a menos que se respalde H_1 con evidencia "más allá de una duda razonable". Sin embargo, en este caso "no rechazar H_0" no implica inocencia, sino sólo que la evidencia fue insuficiente para lograr una condena. Por lo tanto, el jurado no necesariamente *acepta* H_0 sino que *no rechaza* H_0.

10.2 Prueba de una hipótesis estadística

Para ilustrar los conceptos que se utilizan al probar una hipótesis estadística acerca de una población considere el siguiente ejemplo. Se sabe que, después de un periodo de dos años, cierto tipo de vacuna contra un virus que produce resfriado ya sólo es 25% eficaz. Suponga que se eligen 20 personas al azar y se les aplica una vacuna nueva, un poco más costosa, para determinar si protege contra el mismo virus durante un periodo más largo. (En un estudio real de este tipo el número de participantes que reciben la nueva vacuna podría ascender a varios miles. Aquí la muestra es de 20 sólo porque lo único que se busca es demostrar los pasos básicos para realizar una prueba estadística). Si más de 8 individuos de los que reciben la nueva vacuna superan el lapso de 2 años sin contraer el virus, la nueva vacuna se considerará superior a la que se usa en la actualidad. El requisito de que el número exceda a 8 es algo arbitrario, aunque parece razonable, ya que representa una mejoría modesta sobre las 5 personas que se esperaría recibieran protección si fueran inoculadas con la vacuna que actualmente está en uso. En esencia probamos la hipótesis nula de que la nueva vacuna es igual de eficaz después de un periodo de 2 años que la que se utiliza en la actualidad. La hipótesis alternativa es que la nueva vacuna es mejor, y esto equivale a poner a prueba la hipótesis de que el parámetro binomial para la probabilidad de un éxito en un ensayo dado es $p = ¼$, contra la alternativa de que $p > ¼$. Esto por lo general se escribe como se indica a continuación:

$$H_0: p = 0.25,$$
$$H_1: p > 0.25.$$

El estadístico de prueba

El **estadístico de prueba** en el cual se basa nuestra decisión es X, el número de individuos en nuestro grupo de prueba que reciben protección de la nueva vacuna durante un periodo de al menos 2 años. Los valores posibles de X, de 0 a 20, se dividen en dos grupos: los números menores o iguales que 8 y aquellos mayores que 8. Todos los posibles valores mayores que 8 constituyen la **región crítica**. El último número que observamos al pasar a la región crítica se llama **valor crítico**. En nuestro ejemplo el valor crítico es el número 8. Por lo tanto, si $x > 8$, rechazamos H_0 a favor de la hipótesis alternativa H_1. Si $x \leq 8$, no rechazamos H_0. Este criterio de decisión se ilustra en la figura 10.1.

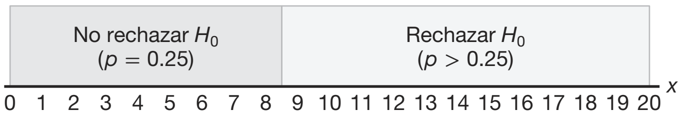

Figura 10.1: Criterio de decisión para probar $p = 0.25$ contra $p > 0.25$.

La probabilidad de un error tipo 1

El procedimiento de toma de decisiones recién descrito podría conducir a cualquiera de dos conclusiones erróneas. Por ejemplo, es probable que la nueva vacuna no sea mejor que la que se usa en la actualidad (H_0 verdadera) y, sin embargo, en este grupo específico de individuos seleccionados aleatoriamente más de 8 pasan el periodo de 2 años sin contraer el virus. Si rechazáramos H_0 a favor de H_1 cuando, de hecho, H_0 es verdadera, cometeríamos un error que se conoce como **error tipo I**.

Definición 10.2: El rechazo de la hipótesis nula cuando es verdadera se denomina **error tipo I**.

Si 8 o menos miembros del grupo superan exitosamente el periodo de 2 años y no concluimos que la nueva vacuna es mejor cuando en realidad sí lo es (H_1 verdadera), cometemos un segundo tipo de error, el de no rechazar la hipótesis H_0 cuando en realidad es falsa. A este error se le conoce como **error tipo II**.

Definición 10.3: No rechazar la hipótesis nula cuando es falsa se denomina **error tipo II**.

Al probar cualquier hipótesis estadística, hay cuatro situaciones posibles que determinan si nuestra decisión es correcta o errónea. Estas cuatro situaciones se resumen en la tabla 10.1.

Tabla 10.1: Situaciones posibles al probar una hipótesis estadística.

	H_0 es verdadera	**H_0 es falsa**
No rechazar H_0	Decisión correcta	Error tipo II
Rechazar H_0	Error tipo I	Decisión correcta

La probabilidad de cometer un error tipo I, también llamada **nivel de significancia**, se denota con la letra griega α. En nuestro ejemplo un error tipo I ocurriría si más de 8 individuos inoculados con la nueva vacuna superan el periodo de 2 años sin contraer el virus y los investigadores concluyen que la nueva vacuna es mejor, cuando en realidad es igual a la vacuna que se utiliza en la actualidad. Por lo tanto, si X es el número de individuos que permanecen sin contraer el virus por al menos dos años,

$$\alpha = P(\text{error tipo I}) = P\left(X > 8 \text{ cuando } p = \frac{1}{4}\right) = \sum_{x=9}^{20} b\left(x; 20, \frac{1}{4}\right)$$

$$= 1 - \sum_{x=0}^{8} b\left(x; 20, \frac{1}{4}\right) = 1 - 0.9591 = 0.0409.$$

Decimos que la hipótesis nula, $p = 1/4$, se prueba al nivel de significancia $\alpha = 0.0409$. En ocasiones el nivel de significancia se conoce como **tamaño de la prueba**. Una región crítica de tamaño 0.0409 es muy pequeña y, por lo tanto, es poco probable que se cometa un error de tipo I. En consecuencia, sería poco probable que más de 8 individuos permanecieran inmunes a un virus durante 2 años utilizando una vacuna nueva que en esencia es equivalente a la que actualmente está en el mercado.

La probabilidad de un error tipo II

La probabilidad de cometer un error tipo II, que se denota con β, es imposible de calcular a menos que tengamos una hipótesis alternativa específica. Si probamos la hipótesis nula $p = 1/4$ contra la hipótesis alternativa $p = 1/2$, entonces podremos calcular la probabilidad de no rechazar H_0 cuando es falsa. Simplemente calculamos la probabilidad de obtener 8 o menos en el grupo que supera el periodo de 2 años cuando $p = 1/2$. En este caso,

$$\beta = P\ (\text{error tipo II}) = P\ \left(X \leq 8 \text{ cuando } p = \frac{1}{2}\right)$$

$$= \sum_{x=0}^{8} b\left(x; 20, \frac{1}{2}\right) = 0.2517.$$

Se trata de una probabilidad elevada que indica un procedimiento de prueba en el cual es muy probable que se rechace la nueva vacuna cuando, de hecho, es mejor a la que está actualmente en uso. De manera ideal, es preferible utilizar un procedimiento de prueba con el cual haya pocas probabilidades de cometer el error tipo I y el error tipo II.

Es posible que el director del programa de prueba esté dispuesto a cometer un error tipo II si la vacuna más costosa no es significativamente mejor. De hecho, la única ocasión en la que desea evitar un error tipo II es cuando el verdadero valor de p es de al menos 0.7. Si $p = 0.7$, este procedimiento de prueba da

$$\beta = P(\text{error tipo II}) = P(X \leq 8 \text{ cuando } p = 0.7)$$
$$= \sum_{x=0}^{8} b(x; 20, 0.7) = 0.0051.$$

Con una probabilidad tan pequeña de cometer un error tipo II es muy improbable que se rechace la nueva vacuna cuando tiene una efectividad de 70% después de un periodo de 2 años. A medida que la hipótesis alternativa se aproxima a la unidad, el valor de β tiende a disminuir hasta cero.

El papel que desempeñan α, β y el tamaño de la muestra

Supongamos que el director del programa de prueba no está dispuesto a cometer un error tipo II cuando la hipótesis alternativa $p = 1/2$ es verdadera, aun cuando se encuentre que la probabilidad de tal error es $\beta = 0.2517$. Siempre es posible reducir β aumentando el tamaño de la región crítica. Por ejemplo, considere lo que les sucede a los valores de α y β cuando cambiamos nuestro valor crítico a 7, de manera que todos los valores mayores que 7 caigan en la región crítica y aquellos menores o iguales que 7 caigan en la región de no rechazo. Así, al probar $p = 1/4$ contra la hipótesis alternativa $p = 1/2$, encontramos que

$$\alpha = \sum_{x=8}^{20} b\left(x; 20, \frac{1}{4}\right) = 1 - \sum_{x=0}^{7} b\left(x; 20, \frac{1}{4}\right) = 1 - 0.8982 = 0.1018$$

$$\beta = \sum_{x=0}^{7} b\left(x; 20, \frac{1}{2}\right) = 0.1316.$$

Al adoptar un nuevo procedimiento de toma de decisiones, reducimos la probabilidad de cometer un error tipo II a costa de aumentar la probabilidad de cometer un error tipo I. Para un tamaño muestral fijo, una disminución en la probabilidad de un error por lo general tendrá como resultado un incremento en la probabilidad del otro error. Por fortuna, **la probabilidad de cometer ambos tipos de errores se puede reducir aumentando el tamaño de la muestra**. Considere el mismo problema usando una muestra aleatoria de 100 individuos. Si más de 36 miembros del grupo superan el periodo de 2 años, rechazamos la hipótesis nula de $p = 1/4$ y aceptamos la hipótesis alternativa de $p > 1/4$. El valor crítico ahora es 36. Todos los valores posibles mayores de 36 constituyen la región crítica y todos los valores posibles menores o iguales que 36 caen en la región de aceptación.

Para determinar la probabilidad de cometer un error tipo I debemos utilizar la aproximación a la curva normal con

$$\mu = np = (100)\left(\frac{1}{4}\right) = 2 \quad \text{y} \quad \sigma = \sqrt{npq} = \sqrt{(100)(1/4)(3/4)} = 4.33.$$

Con respecto a la figura 10.2, necesitamos el área bajo la curva normal a la derecha de $x = 36.5$. El valor z correspondiente es

$$z = \frac{36.5 - 25}{4.33} = 2.66.$$

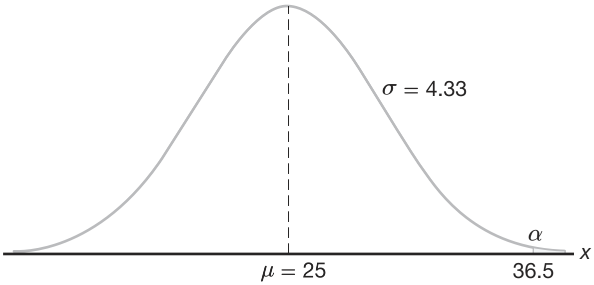

Figura 10.2: Probabilidad de un error tipo I.

En la tabla A.3 encontramos que

$$\alpha = P(\text{error tipo I}) = P\left(X > 36 \text{ cuando } p = \frac{1}{4}\right) \approx P(Z > 2.66)$$
$$= 1 - P(Z < 2.66) = 1 - 0.9961 = 0.0039.$$

Si H_0 es falsa y el verdadero valor de H_1 es $p = 1/2$, determinamos la probabilidad de un error tipo II usando la aproximación a la curva normal con

$$\mu = np = (100)(1/2) = 50 \quad \text{y} \quad \sigma = \sqrt{npq} = \sqrt{(100)(1/2)(1/2)} = 5 .$$

La probabilidad de que un valor caiga en la región de no rechazo cuando H_0 es verdadera es dada por el área de la región sombreada a la izquierda de $x = 36.5$ en la figura 10.3. El valor z que corresponde a $x = 36.5$ es

$$z = \frac{36.5 - 50}{5} = -2.7.$$

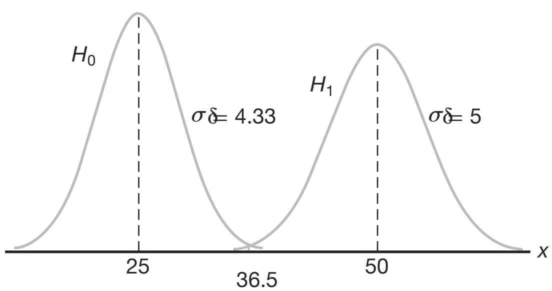

Figura 10.3: Probabilidad de un error tipo II.

Por lo tanto,

$$\beta = P(\text{error tipo II}) = P\left(X \leq 36 \text{ cuando } p = \frac{1}{2}\right) \approx P(Z < -2.7) = 0.0035.$$

Evidentemente, los errores tipo I y tipo II rara vez ocurren si el experimento consta de 100 individuos.

El ejemplo anterior destaca la estrategia del científico en la prueba de hipótesis. Después de que se plantean las hipótesis nula y alternativa es importante considerar la sensibilidad del procedimiento de prueba. Con esto queremos decir que debería determinarse un valor razonable a una α fija para la probabilidad de aceptar de manera errónea H_0, es decir, el valor de β, cuando la verdadera situación representa alguna *desviación importante de* H_0. Por lo general, es posible determinar un valor para el tamaño de la muestra, para el que existe un equilibrio razonable entre los valores de α y β que se calcula de esta manera. El problema de la vacuna es un ejemplo.

Ilustración con una variable aleatoria continua

Los conceptos que se analizan aquí para una población discreta también se pueden aplicar a variables aleatorias continuas. Considere la hipótesis nula de que el peso promedio de estudiantes hombres en cierta universidad es de 68 kilogramos, contra la hipótesis alternativa de que es diferente a 68. Es decir, deseamos probar

$$H_0\colon \mu = 68,$$
$$H_1\colon \mu \neq 68.$$

La hipótesis alternativa nos permite la posibilidad de que $\mu < 68$ o $\mu > 68$.

Una media muestral que caiga cerca del valor hipotético de 68 se consideraría como evidencia a favor de H_0. Por otro lado, una media muestral considerablemente menor que o mayor que 68 sería evidencia en contra de H_0 y, por lo tanto, favorecería a H_1. La media muestral es el estadístico de prueba en este caso. Una región crítica para el estadístico de prueba se puede elegir de manera arbitraria como los dos intervalos $\bar{x} < 67$ y $\bar{x} > 69$. La región de no rechazo será entonces el intervalo $67 \leq \bar{x} \leq 69$. Este criterio de decisión se ilustra en la figura 10.4.

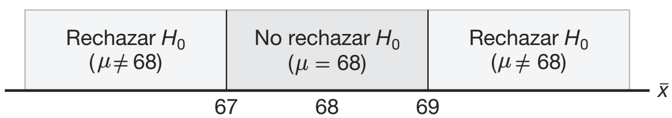

Figura 10.4: Región crítica (en azul).

Utilicemos ahora el criterio de decisión de la figura 10.4 para calcular las probabilidades de cometer los errores tipo I y tipo II cuando probemos la hipótesis nula $\mu = 68$ kilogramos contra la alternativa $\mu \neq 68$ kilogramos.

Suponga que la desviación estándar de la población de pesos es $\sigma = 3.6$. Para muestras grandes podemos sustituir s por σ si no disponemos de ninguna otra estimación de σ. Nuestro estadístico de decisión, que se basa en una muestra aleatoria de tamaño $n = 36$, será $\bar{X}$, el estimador más eficaz de μ. Del teorema del límite central sabemos que la distribución muestral de $\bar{X}$ es aproximadamente normal con desviación estándar $\sigma_{\bar{X}} = \sigma/\sqrt{n} = 3.6/6 = 0.6$.

La probabilidad de cometer un error tipo I, o el nivel de significancia de nuestra prueba, es igual a la suma de las áreas sombreadas en cada cola de la distribución en la figura 10.5. Por lo tanto,

$$\alpha = P(\bar{X} < 67 \text{ cuando } \mu = 68) + P(\bar{X} > 69 \text{ cuando } \mu = 68).$$

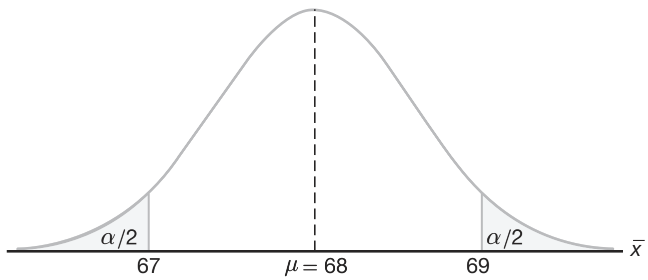

Figura 10.5: Región crítica para probar $\mu = 68$ contra $\mu \neq 68$.

Los valores z correspondientes a $\bar{x}_1 = 67$ y $\bar{x}_2 = 69$ cuando H_0 es verdadera son

$$z_1 = \frac{67 - 68}{0.6} = -1.67 \quad \text{y} \quad z_2 = \frac{69 - 68}{0.6} = 1.67.$$

Por lo tanto,

$$\alpha = P(Z < -1.67) + P(Z > 1.67) = 2P(Z < -1.67) = 0.0950.$$

Por consiguiente, 9.5% de todas las muestras de tamaño 36 nos conducirían a rechazar $\mu = 68$ kilogramos cuando, de hecho, ésta es verdadera. Para reducir α tenemos que elegir entre aumentar el tamaño de la muestra o ampliar la región de no rechazo. Suponga que aumentamos el tamaño de la muestra a $n = 64$. Entonces $\sigma_{\bar{X}} = 3.6/8 = 0.45$. En consecuencia,

$$z_1 = \frac{67 - 68}{0.45} = -2.22 \quad \text{y} \quad z_2 = \frac{69 - 68}{0.45} = 2.22.$$

Por lo tanto,

$$\alpha = P(Z < -2.22) + P(Z > 2.22) = 2\, P(Z < -2.22) = 0\ .0264.$$

La reducción de α no es suficiente por sí misma para garantizar un buen procedimiento de prueba. Debemos evaluar β para varias hipótesis alternativas. Si es importante rechazar H_0 cuando la media verdadera sea algún valor $\mu \geq 70$ o $\mu \leq 66$, entonces se debería calcular y examinar la probabilidad de cometer un error tipo II para las alternativas $\mu = 66$ y $\mu = 70$. Debido a la simetría, sólo es necesario considerar la probabilidad de no rechazar la hipótesis nula $\mu = 68$ cuando la alternativa $\mu = 70$ es verdadera. Cuando la media muestral $\bar{x}$ caiga entre 67 y 69, cuando H_1 sea verdadera, resultará un error tipo II. Por lo tanto, remitiéndonos a la figura 10.6 encontramos que

$$\beta = P(67 \leq \bar{X} \leq 69 \text{ cuando } \mu = 70).$$

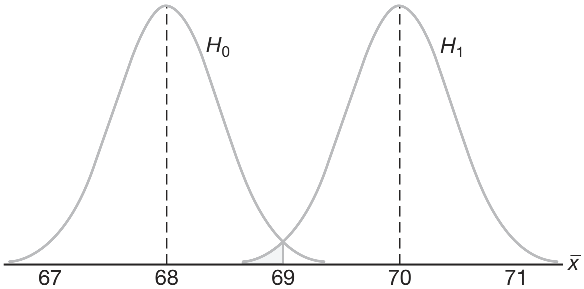

Figura 10.6: Probabilidad del error tipo II al probar $\mu = 68$ contra $\mu = 70$.

Los valores z que corresponden a $\bar{x}_1 = 67$ y $\bar{x}_2 = 69$ cuando H_1 es verdadera son

$$z_1 = \frac{67 - 70}{0.45} = -6.67 \quad \text{y} \quad z_2 = \frac{69 - 70}{0.45} = -2.22.$$

Por lo tanto,

$$\beta = P(-6.67 < Z < -2.22) = P(Z < -2.22) - P(Z < -6.67)$$
$$= 0.0132 - 0.0000 = 0.0132.$$

Si el valor verdadero de μ es la alternativa $\mu = 66$, el valor de β nuevamente será 0.0132. Para todos los valores posibles de $\mu < 66$ o $\mu > 70$, el valor de β será incluso más pequeño cuando $n = 64$ y, en consecuencia, habrá poca oportunidad de no rechazar H_0 cuando sea falsa.

La probabilidad de cometer un error tipo II aumenta rápidamente cuando el valor verdadero de μ se aproxima al valor hipotético pero no es igual a éste. Desde luego, ésta suele ser la situación en la que no nos importa cometer un error tipo II. Por ejemplo, si la hipótesis alternativa $\mu = 68.5$ es verdadera, no nos importa cometer un error tipo II al concluir que la respuesta verdadera es $\mu = 68$. La probabilidad de cometer tal error será elevada cuando $n = 64$. Al remitirnos a la figura 10.7, tenemos

$$\beta = P(67 \leq \bar{X} \leq 69 \text{ cuando } \mu = 68.5).$$

Los valores z correspondientes a $\bar{x}_1 = 67$ y $\bar{x}_2 = 69$ cuando $\mu = 68.5$ son

$$z_1 = \frac{67 - 68.5}{0.45} = -3.33 \quad \text{y} \quad z_2 = \frac{69 - 68.5}{0.45} = 1.11.$$

Por lo tanto,

$$\beta = P(-3.33 < Z < 1.11) = P(Z < 1.11) - P(Z < -3.33)$$
$$= 0.8665 - 0.0004 = 0.8661.$$

Los ejemplos anteriores ilustran las siguientes propiedades importantes:

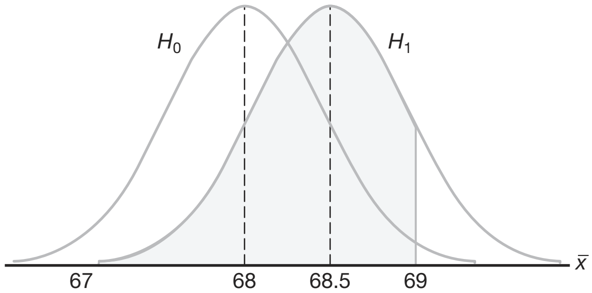

Figura 10.7: Error tipo II para la prueba de $\mu = 68$ contra $\mu = 68.5$.

Propiedades importantes de una prueba de hipótesis

1. Los errores tipo I y tipo II están relacionados. Por lo general una disminución en la probabilidad de cometer uno da como resultado un incremento en la probabilidad de cometer el otro.
2. El tamaño de la región crítica y, por lo tanto, la probabilidad de cometer un error tipo I, siempre se puede reducir ajustando el (los) valor(es) crítico(s).
3. Un aumento en el tamaño de la muestra n reducirá α y β de forma simultánea.
4. Si la hipótesis nula es falsa, β es un máximo cuando el valor verdadero de un parámetro se aproxima al valor hipotético. Cuanto más grande sea la distancia entre el valor verdadero y el valor hipotético, más pequeña será β.

Definición 10.4: La **potencia** de una prueba es la probabilidad de rechazar H_0 dado que una alternativa específica es verdadera.

La potencia de una prueba se puede calcular como $1 - \beta$. A menudo **diferentes tipos de pruebas se comparan contrastando propiedades de potencia**. Considere el caso anterior en el que probamos H_0: $\mu = 68$ y H_1: $\mu \neq 68$. Como antes, suponga que nos interesa evaluar la sensibilidad de la prueba, la cual es determinada por la regla de que no rechazamos H_0 si $67 \leq \bar{x} \leq 69$. Buscamos la capacidad de la prueba para rechazar H_0 de manera adecuada cuando en realidad $\mu = 68.5$. Vimos que la probabilidad de un error tipo II es dada por $\beta = 0.8661$. Por consiguiente, la **potencia** de la prueba es $1 - 0.8661 = 0.1339$. En cierto sentido, la potencia es una medida más sucinta de cuán sensible es la prueba para detectar diferencias entre una media de 68 y otra de 68.5. En este caso, si μ es verdaderamente 68.5, la prueba como se describe *rechazará de forma adecuada H_0 sólo 13.39% de las veces.* Como resultado, la prueba no sería buena si es importante que el analista tenga una oportunidad razonable de distinguir realmente entre una media de 68.0 (que especifica H_0) y una media de 68.5. De lo anterior resulta claro que para producir una potencia deseable, digamos, mayor que 0.8, es necesario incrementar α o aumentar el tamaño de la muestra.

Hasta ahora gran parte del análisis de la prueba de hipótesis se ha enfocado en los principios y las definiciones. En las secciones que siguen seremos más específicos y clasificaremos las hipótesis en categorías. También estudiaremos pruebas de hipótesis sobre varios parámetros de interés. Comenzamos estableciendo la diferencia entre hipótesis unilaterales y bilaterales.

Pruebas de una y dos colas

Una prueba de cualquier hipótesis estadística donde la alternativa es **unilateral**, como

$$H_0\colon \theta = \theta_0,$$
$$H_1\colon \theta > \theta_0,$$

o quizás

$$H_0\colon \theta = \theta_0,$$
$$H_1\colon \theta < \theta_0,$$

se denomina **prueba de una sola cola**. Anteriormente en esta sección se hizo referencia al **estadístico de prueba** para una hipótesis. Por lo general la región crítica para la hipótesis alternativa $\theta > \theta_0$ yace en la cola derecha de la distribución del estadístico de prueba, en tanto que la región crítica para la hipótesis alternativa $\theta < \theta_0$ yace por completo en la cola izquierda. (En cierto sentido el símbolo de desigualdad señala la dirección en donde se encuentra la región crítica). En el experimento de la vacuna se utilizó una prueba de una sola cola para probar la hipótesis $p = 1/4$ contra la alternativa unilateral $p > 1/4$ para la distribución binomial. La región crítica de una sola cola por lo general es evidente; el lector debería visualizar el comportamiento del estadístico de prueba y observar la *señal* evidente que produciría evidencia que respalde la hipótesis alternativa.

La prueba de cualquier hipótesis alternativa donde la alternativa es **bilateral**, como

$$H_0\colon \theta = \theta_0,$$
$$H_1\colon \theta \neq \theta_0,$$

se denomina **prueba de dos colas**, ya que la región crítica se divide en dos partes, a menudo con probabilidades iguales en cada cola de la distribución del estadístico de prueba. La hipótesis alternativa $\theta \neq \theta_0$ establece que $\theta < \theta_0$ o que $\theta > \theta_0$. Se utilizó una prueba de dos colas para probar la hipótesis nula $\mu = 68$ kilogramos contra la alternativa bilateral $\mu \neq 68$ kilogramos en el ejemplo de la población continua de los pesos de estudiantes.

¿Cómo se eligen las hipótesis nula y alternativa?

Con frecuencia la hipótesis nula H_0 se plantea usando el *signo de igualdad.* Con este método se observa claramente cómo se controla la probabilidad de cometer un error tipo I. Sin embargo, hay situaciones en que "no rechazar H_0" implica que el parámetro θ podría ser cualquier valor definido por el complemento natural de la hipótesis alternativa. Por ejemplo, en el caso de la vacuna, donde la hipótesis alternativa es $H_1\colon p > 1/4$, es muy posible que el no rechazo de H_0 no pueda descartar un valor de p menor que 1/4. Sin embargo, es evidente que en el caso de las pruebas de una cola la consideración más importante es el planteamiento de la alternativa.

La decisión de plantear una prueba de una cola o una de dos colas depende de la conclusión que se obtenga si se rechaza H_0. La ubicación de la región crítica sólo se puede determinar después de que se plantea H_1. Por ejemplo, al probar una medicina nueva se establece la hipótesis de que no es mejor que las medicinas similares que actualmente hay en el mercado y se prueba contra la hipótesis alternativa de que la medicina nueva es mejor. Esta hipótesis alternativa dará como resultado una prueba de una sola cola, con la región crítica en la cola derecha. Sin embargo, si deseamos comparar una nueva técnica de enseñanza con el procedimiento convencional del salón de clases, la hipótesis alternativa debe permitir que el nuevo método sea inferior o superior al procedimiento convencional. Por lo tanto, la prueba sería de dos colas con la región crítica dividida en partes iguales, de manera que caiga en los extremos de las colas izquierda y derecha de la distribución de nuestro estadístico.

Ejemplo 10.1: Un fabricante de cierta marca de cereal de arroz afirma que el contenido promedio de grasa saturada no excede a 1.5 gramos por porción. Plantee las hipótesis nula y alternativa que se utilizarán para probar esta afirmación y establezca en dónde se localiza la región crítica.

***Solución*:** La afirmación del fabricante se rechazará sólo si μ es mayor que 1.5 miligramos y no se rechazará si μ es menor o igual que 1.5 miligramos. Entonces, probamos

$$H_0\colon \mu = 1.5,$$
$$H_1\colon \mu > 1.5.$$

El hecho de no rechazar H_0 no descarta valores menores que 1.5 miligramos. Como tenemos una prueba de una cola, el símbolo mayor indica que la región crítica reside por completo en la cola derecha de la distribución de nuestro estadístico de prueba $\bar{X}$. ■

Ejemplo 10.2: Un agente de bienes raíces afirma que 60% de todas las viviendas privadas que se construyen actualmente son casas con tres dormitorios. Para probar esta afirmación se inspecciona una muestra grande de viviendas nuevas. Se registra la proporción de las casas con 3 dormitorios y se utiliza como estadístico de prueba. Plantee las hipótesis nula y alternativa que se utilizarán en esta prueba y determine la ubicación de la región crítica.

***Solución*:** Si el estadístico de prueba fuera considerablemente mayor o menor que $p = 0.6$, rechazaríamos la afirmación del agente. En consecuencia, deberíamos plantear las siguientes hipótesis:

$$H_0\colon p = 0.6,$$
$$H_1\colon p \neq 0.6.$$

La hipótesis alternativa implica una prueba de dos colas con la región crítica dividida por igual en ambas colas de la distribución de $\hat{P}$, nuestro estadístico de prueba. ■

10.3 Uso de valores *P* para la toma de decisiones en la prueba de hipótesis

Al probar hipótesis en las que el estadístico de prueba es discreto, la región crítica se podría elegir de manera arbitraria y determinar su tamaño. Si α es demasiado grande, se reduce haciendo un ajuste en el valor crítico. Quizá sea necesario aumentar el tamaño de la muestra para compensar la disminución que ocurre de manera automática en la potencia de la prueba.

Por generaciones enteras de análisis estadístico se ha vuelto costumbre elegir una α de 0.05 o 0.01 y seleccionar la región crítica de acuerdo con esto. Entonces, desde luego, el rechazo o no rechazo estrictos de H_0 dependerá de esa región crítica. Por ejemplo, si la prueba es de dos colas, α se fija a un nivel de significancia de 0.05 y el estadístico de prueba implica, digamos, la distribución normal estándar, entonces se observa un valor z de los datos y la región crítica es

$$z > 1.96 \quad \text{o} \quad z < -1.96,$$

donde el valor 1.96 corresponde a $z_{0.025}$ en la tabla A.3. Un valor de z en la región crítica sugiere la aseveración: "El valor del estadístico de prueba es significativo", el cual se puede traducir al lenguaje del caso. Por ejemplo, si la hipótesis es dada por

$$H_0: \mu = 10,$$
$$H_1: \mu \neq 10,$$

se puede decir: "La media difiere de manera significativa del valor 10".

Preselección de un nivel de significancia

Esta preselección de un nivel de significancia α tiene sus raíces en la filosofía de que se debe controlar el riesgo máximo de cometer un error tipo I. Sin embargo, este enfoque no explica los valores del estadístico de prueba que están "cercanos" a la región crítica. Suponga, por ejemplo, que en el caso de H_0: $\mu = 10$, contra H_1: $\mu \neq 10$, se observa un valor $z = 1.87$. En términos estrictos, con $\alpha = 0.05$ el valor no es significativo; pero el riesgo de cometer un error tipo I si se rechaza H_0 en este caso difícilmente se podría considerar grave. De hecho, en una situación de dos colas, el riesgo se cuantifica como

$$P = 2P(Z > 1.87 \text{ cuando } \mu = 10) = 2(0.0307) = 0.0614.$$

Como resultado, 0.0614 es la probabilidad de obtener un valor de z tan grande o mayor (en magnitud) que 1.87 cuando, de hecho, $\mu = 10$. Aunque esta evidencia en contra de H_0 no es tan firme como la que resultaría de un rechazo a un nivel $\alpha = 0.05$, se trata de información importante para el usuario. De hecho, el uso continuo de $\alpha = 0.05$ o 0.01 tan sólo es un resultado de lo que los estándares han transmitido por generaciones. **En la estadística aplicada los usuarios han adoptado de forma extensa el método del valor *P***. El método está diseñado para dar al usuario una alternativa (en términos de una probabilidad) a la mera conclusión de "rechazo" o "no rechazo". El cálculo del valor P también proporciona al usuario información importante cuando el valor z cae *dentro de la región crítica ordinaria.* Por ejemplo, si z es 2.73, resulta informativo para el usuario observar que

$$P = 2(0.0032) = 0.0064,$$

y, por consiguiente, el valor z es significativo a un nivel considerablemente menor que 0.05. Es importante saber que bajo la condición de H_0 un valor de $z = 2.73$ es un evento demasiado raro. A saber, un valor al menos tan grande en magnitud sólo ocurriría 64 veces en 10,000 experimentos.

Demostración gráfica de un valor *P*

Una manera muy simple de explicar gráficamente un valor P consiste en considerar dos muestras distintas. Suponga que se están considerando dos materiales para cubrir un tipo específico de metal con el fin de evitar la corrosión. Se obtienen especímenes y se cubre un grupo con el material 1 y otro grupo con el material 2. Los tamaños muestrales son $n_1 = n_2 = 10$ para cada muestra y la corrosión se mide en el porcentaje del área superficial afectada. La hipótesis plantea que las muestras provienen de distribuciones comunes con media $\mu = 10$. Supongamos que la varianza de la población es 1.0. Entonces, probamos

$$H_0: \mu_1 = \mu_2 = 10.$$

Representemos con la figura 10.8 una gráfica de puntos de los datos. Los datos se colocan en la distribución determinada por la hipótesis nula. Supongamos que los datos "×" se refieren al material 1 y que los datos "∘" se refieren al material 2. Parece evidente que los datos realmente refutan la hipótesis nula. Pero, ¿cómo se podría resumir esto en un número? **El valor *P* se puede considerar simplemente como la probabilidad de obtener este conjunto de datos dado que las muestras provienen de la misma distribución**. Es evidente que esta probabilidad es muy pequeña, ¡digamos 0.00000001! Por consiguiente, el pequeño valor P evidentemente refuta H_0, y la conclusión es que las medias de la población son significativamente diferentes.

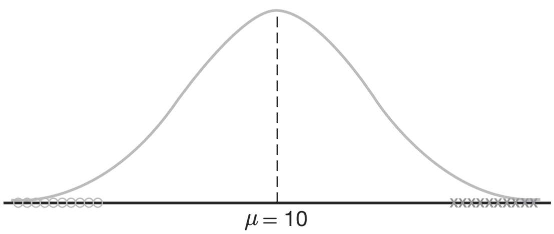

Figura 10.8: Datos que son probablemente generados de poblaciones que tienen dos medias diferentes.

El uso del método del valor P como auxiliar en la toma de decisiones es muy natural y casi todos los programas de cómputo que proporcionan el cálculo de pruebas de hipótesis ofrecen valores P, junto con valores del estadístico de prueba adecuado. La siguiente es una definición formal de un valor P.

Definición 10.5: Un **valor *P*** es el nivel (de significancia) más bajo en el que el valor observado del estadístico de prueba es significativo.

¿En qué difiere el uso de los valores *P* de la prueba de hipótesis clásica?

En este momento resulta tentador resumir los procedimientos que se asocian con la prueba de, digamos, H_0: $\theta = \theta_0$. Sin embargo, el estudiante que es novato en esta área deberá tener en cuenta que hay diferencias entre el enfoque y la filosofía del método clásico de α fija, que tiene su momento más importante en la conclusión de “rechazar H_0” o “no rechazar H_0” y el método del valor P. En este último no se determina una α fija y las conclusiones se obtienen con base en el tamaño del valor P, según la apreciación subjetiva del ingeniero o del científico. Aun cuando los modernos programas de cómputo proporcionan valores P, es importante que el lector comprenda ambos enfoques para apreciar la totalidad de los conceptos. Por lo tanto, ofrecemos una breve lista con los pasos del procedimiento tanto para el método clásico como para el del valor P.

Aproximación a la prueba de hipótesis con probabilidad fija del error tipo I

1. Establezca las hipótesis nula y alternativa.
2. Elija un nivel de significancia α fijo.
3. Seleccione un estadístico de prueba adecuado y establezca la región crítica con base en α.
4. Rechace H_0 si el estadístico de prueba calculado está en la región crítica. De otra manera, no rechace H_0.
5. Saque conclusiones científicas y de ingeniería.

Prueba de significancia (método del valor P)

1. Establezca las hipótesis nula y alternativa.
2. Elija un estadístico de prueba adecuado.
3. Calcule el valor P con base en los valores calculados del estadístico de prueba.
4. Saque conclusiones con base en el valor P y los conocimientos del sistema científico.

En secciones posteriores de este capítulo y en los capítulos siguientes muchos ejemplos y ejercicios destacarán el método del valor P para obtener conclusiones científicas.

Ejercicios

10.1 Suponga que un alergólogo desea probar la hipótesis de que al menos 30% del público es alérgico a algunos productos de queso. Explique cómo el alergólogo podría cometer
a) un error tipo I;
b) un error tipo II.

10.2 Una socióloga se interesa en la eficacia de un curso de entrenamiento diseñado para lograr que más conductores utilicen los cinturones de seguridad en los automóviles.
a) ¿Qué hipótesis pone a prueba si comete un error tipo I al concluir de manera errónea que el curso de entrenamiento no es eficaz?
b) ¿Qué hipótesis pone a prueba si comete un error tipo II al concluir de forma errónea que el curso de entrenamiento es eficaz?

10.3 Se acusa a una empresa grande de discriminación en sus prácticas de contratación.
a) ¿Qué hipótesis se pone a prueba si un jurado comete un error tipo I al encontrar culpable a la empresa?
b) ¿Qué hipótesis se pone a prueba si un jurado comete un error tipo II al encontrar culpable a la empresa?

10.4 Un fabricante de telas considera que la proporción de pedidos de materia prima que llegan con retraso es $p = 0.6$. Si una muestra aleatoria de 10 pedidos indica que 3 o menos llegaron con retraso, la hipótesis de que $p = 0.6$ se debería rechazar a favor de la alternativa $p < 0.6$. Utilice la distribución binomial.
a) Calcule la probabilidad de cometer un error tipo I si la proporción verdadera es $p = 0.6$.
b) Calcule la probabilidad de cometer un error tipo II para las alternativas $p = 0.3$, $p = 0.4$ y $p = 0.5$.

10.5 Repita el ejercicio 10.4 pero suponga que se seleccionan 50 pedidos y que se define a la región crítica como $x \leq 24$, donde x es el número de pedidos en la muestra que llegaron con retraso. Utilice la aproximación normal.

10.6 Se estima que la proporción de adultos que vive en una pequeña ciudad que son graduados universitarios es $p = 0.6$. Para probar esta hipótesis se selecciona una muestra aleatoria de 15 adultos. Si el número de graduados en la muestra es cualquier número entre 6 y 12, no rechazaremos la hipótesis nula de que $p = 0.6$; de otro modo, concluiremos que $p \neq 0.6$.
a) Evalúe α suponiendo que $p = 0.6$. Utilice la distribución binomial.

b) Evalúe β para las alternativas $p = 0.5$ y $p = 0.7$.
c) ¿Es éste un buen procedimiento de prueba?

10.7 Repita el ejercicio 10.6 pero suponga que se seleccionan 200 adultos y que la región de no rechazo se define como $110 \leq x \leq 130$, donde x es el número de individuos graduados universitarios en la muestra. Utilice la aproximación normal.

10.8 En la publicación *Relief from Arthritis* de Thorsons Publishers, Ltd., John E. Croft afirma que más de 40% de los individuos que sufren de osteoartritis experimentan un alivio medible con un ingrediente producido por una especie particular de mejillón que se encuentra en la costa de Nueva Zelanda. Para probar esa afirmación se suministra el extracto de mejillón a un grupo de 7 pacientes con osteoartritis. Si 3 o más de los pacientes experimentan alivio, no rechazaremos la hipótesis nula de que $p = 0.4$; de otro modo, concluiremos que $p < 0.4$.
a) Evalúe α suponiendo que $p = 0.4$.
b) Evalúe β para la alternativa $p = 0.3$.

10.9 Una tintorería afirma que un nuevo removedor de manchas quitará más de 70% de las manchas en las que se aplique. Para verificar esta afirmación el removedor de manchas se utilizará sobre 12 manchas elegidas al azar. Si se eliminan menos de 11 de las manchas, no se rechazará la hipótesis nula de que $p = 0.7$; de otra manera, concluiremos que $p > 0.7$.
a) Evalúe α, suponiendo que $p = 0.7$.
b) Evalúe β para la alternativa $p = 0.9$.

10.10 Repita el ejercicio 10.9 pero suponga que se tratan 100 manchas y que la región crítica se define como $x > 82$, donde x es el número de manchas eliminadas.

10.11 Repita el ejercicio 10.8 pero suponga que el extracto de mejillón se administra a 70 pacientes y que la región crítica se define como $x < 24$, donde x es el número de pacientes con osteoartritis que experimentan alivio.

10.12 Se pregunta a una muestra aleatoria de 400 votantes en cierta ciudad si están a favor de un impuesto adicional de 4% sobre las ventas de gasolina con el fin de obtener los fondos que se necesitan con urgencia para la reparación de calles. Si más de 220 votantes, pero menos de 260 de ellos, favorecen el impuesto sobre las ventas, concluiremos que 60% de los votantes lo apoyan.
a) Calcule la probabilidad de cometer un error tipo I si 60% de los votantes están a favor del aumento de impuestos.
b) ¿Cuál es la probabilidad de cometer un error tipo II al utilizar este procedimiento de prueba si en realidad sólo 48% de los votantes está a favor del impuesto adicional a la gasolina?

10.13 Suponga que en el ejercicio 10.12 concluimos que 60% de los votantes está a favor del impuesto sobre las ventas de gasolina si más de 214 votantes, pero menos de 266 de ellos, lo favorecen. Demuestre que esta nueva región crítica tiene como resultado un valor más pequeño para α a costa de aumentar β.

10.14 Un fabricante desarrolla un nuevo sedal para pesca que, según afirma, tiene una resistencia media a la rotura de 15 kilogramos con una desviación estándar de 0.5 kilogramos. Para probar la hipótesis de que $\mu =$ 15 kilogramos contra la alternativa de que $\mu < 15$ kilogramos se prueba una muestra aleatoria de 50 sedales. La región crítica se define como $\bar{x} < 14.9$.
a) Calcule la probabilidad de cometer un error tipo I cuando H_0 es verdadera.
b) Evalúe β para las alternativas $\mu = 14.8$ y $\mu = 14.9$ kilogramos.

10.15 En un restaurante de carnes una máquina de bebidas gaseosas se ajusta para que la cantidad de bebida que sirva se distribuya de forma aproximadamente normal, con una media de 200 mililitros y una desviación estándar de 15 mililitros. La máquina se verifica periódicamente tomando una muestra de 9 bebidas y calculando el contenido promedio. Si $\bar{x}$ cae en el intervalo $191 < \bar{x} < 209$, se considera que la máquina opera de forma satisfactoria; de otro modo, se concluye que $\mu \neq 200$ mililitros.
a) Calcule la probabilidad de cometer un error tipo I cuando $\mu = 200$ mililitros.
b) Calcule la probabilidad de cometer un error tipo II cuando $\mu = 215$ mililitros.

10.16 Repita el ejercicio 10.15 para muestras de tamaño $n = 25$. Utilice la misma región crítica.

10.17 Se desarrolla un nuevo proceso de cura para cierto tipo de cemento que da como resultado una resistencia media a la compresión de 5000 kilogramos por centímetro cuadrado y una desviación estándar de 120 kilogramos. Para probar la hipótesis de que $\mu =$ 5000 contra la alternativa de que $\mu < 5000$ se toma una muestra aleatoria de 50 piezas de cemento. La región crítica se define como $\bar{x} < 4970$.
a) Calcule la probabilidad de cometer un error tipo I cuando H_0 es verdadera.
b) Evalúe β para las alternativas $\mu = 4970$ y $\mu =$ 4960.

10.18 Si graficamos las probabilidades de no rechazar H_0 que corresponden a diversas alternativas para μ (incluido el valor especificado para H_0) y conectamos todos los puntos mediante una curva suave, obtenemos la **curva característica de operación** del criterio de prueba o, simplemente, la curva CO. Observe que la probabilidad de no rechazar H_0 cuando es verdadera es simplemente $1 - \alpha$. Las curvas características de operación se utilizan con amplitud en aplicaciones industriales para proporcionar una muestra visual de los méritos del criterio de prueba. Remítase al ejercicio 10.15 y calcule las probabilidades de no rechazar H_0 para los siguientes 9 valores de μ y grafique la curva CO: 184, 188, 192, 196, 200, 204, 208, 212 y 216.

10.4 Una sola muestra: pruebas respecto a una sola media

En esta sección consideramos de manera formal pruebas de hipótesis para una sola media de la población. Muchos de los ejemplos de las secciones anteriores incluyen pruebas sobre la media, por lo que el lector ya debería tener una idea de algunos de los detalles que aquí se describen.

Pruebas para una sola media (varianza conocida)

Primero deberíamos describir las suposiciones en las que se basa el experimento. El modelo para la situación subyacente se centra alrededor de un experimento con $X_1, X_2, \ldots, X_n$, que representan una muestra aleatoria de una distribución con media μ y varianza $\sigma^2 > 0$. Considere primero la hipótesis

$$H_0: \mu = \mu_0,$$
$$H_1: \mu \neq \mu_0.$$

El estadístico de prueba adecuado se debe basar en la variable aleatoria $\bar{X}$. En el capítulo 8 se presentó el teorema del límite central, el cual establece en esencia que, sin importar la distribución de X, la variable aleatoria $\bar{X}$ tiene una distribución casi normal con media μ y varianza σ^2/n para muestras de tamaño razonablemente grande. Por consiguiente, $\mu_{\bar{X}} = \mu$ y $\sigma_{\bar{X}}^2 = \sigma^2/n$. Podemos determinar, entonces, una región crítica basada en el promedio muestral calculado $\bar{x}$. Ahora ya debería quedarle claro al lector que habrá una región crítica de dos colas para la prueba.

Estandarización de $\bar{X}$

Es conveniente estandarizar $\bar{X}$ e incluir de manera formal la variable aleatoria **normal estándar** Z, donde

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}.$$

Sabemos que, *bajo* H_0, es decir, si $\mu = \mu_0$, entonces $\sqrt{n}\left(\bar{X} - \mu_0\right)/\sigma$ tiene una distribución $n(x; 0, 1)$ y, por lo tanto, la expresión

$$P\left(-z_{\alpha/2} < \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} < z_{\alpha/2}\right) = 1 - \alpha$$

se puede utilizar para escribir una región de no rechazo adecuada. El lector debería tener en la mente que, formalmente, la región crítica se diseña para controlar α, la probabilidad de cometer un error tipo I. Debería ser evidente que se necesita una *señal de evidencia de dos colas* para apoyar H_1. Así, dado un valor calculado $\bar{x}$, la prueba formal implica rechazar H_0 si el *estadístico de prueba* z calculado cae en la región crítica que se describe a continuación.

Procedimiento de prueba para una sola media (varianza conocida)

$$z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} > z_{\alpha/2} \quad \text{o} \quad z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} < -z_{\alpha/2}$$

Si $-z_{\alpha/2} < z < z_{\alpha/2}$, no se rechaza H_0. El rechazo de H_0, desde luego, implica la aceptación de la hipótesis alternativa $\mu \neq \mu_0$. Con esta definición de la región crítica debería quedar claro que habrá α probabilidades de rechazar H_0 (al caer en la región crítica) cuando, en realidad, $\mu = \mu_0$.

Aunque es más fácil entender la región crítica escrita en términos de z, escribimos la misma región crítica en términos del promedio calculado $\bar{x}$. Lo siguiente se puede escribir como un procedimiento de decisión idéntico:

$$\textit{rechazar } H_0 \text{ si } \bar{x} < a \text{ o } \bar{x} > b,$$

donde

$$a = \mu_0 - z_{\alpha/2}\frac{\sigma}{\sqrt{n}}, \qquad b = \mu_0 + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}.$$

En consecuencia, para un nivel de significancia α, los valores críticos de la variable aleatoria z y $\bar{x}$ se presentan en la figura 10.9.

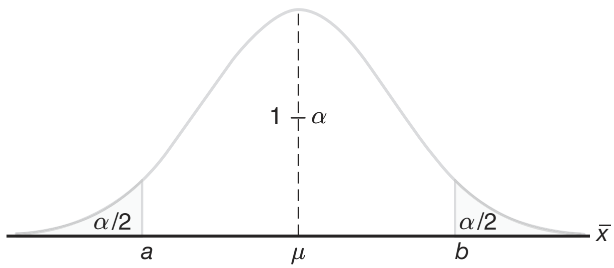

Figura 10.9: Región crítica para la hipótesis alternativa $\mu \neq \mu_0$.

Las pruebas de hipótesis unilaterales sobre la media incluyen el mismo estadístico que se describe en el caso bilateral. La diferencia, por supuesto, es que la región crítica sólo está en una cola de la distribución normal estándar. Por ejemplo, supongamos que buscamos probar

$$H_0: \mu = \mu_0,$$

$$H_1: \mu > \mu_0.$$

La señal que favorece H_1 proviene de *valores grandes* de z. Así, el rechazo de H_0 resulta cuando se calcula $z > z_\alpha$. Evidentemente, si la alternativa es H_1: $\mu < \mu_0$, la región crítica está por completo en la cola inferior, por lo que el rechazo resulta de $z < -z_\alpha$. Aunque en el caso de una prueba unilateral la hipótesis nula se puede escribir como H_0: $\mu \leq \mu_0$ o H_0: $\mu \geq \mu_0$, por lo general se escribe como H_0: $\mu = \mu_0$.

Los siguientes dos ejemplos ilustran pruebas de medias para el caso en el que se conoce σ.

Ejemplo 10.3: Una muestra aleatoria de 100 muertes registradas en Estados Unidos el año pasado reveló una vida promedio de 71.8 años. Si se supone una desviación estándar de la población de 8.9 años, ¿esto parece indicar que la vida media actual es mayor que 70 años? Utilice un nivel de significancia de 0.05.

***Solución*:**
1. H_0: $\mu = 70$ años.
2. H_1: $\mu > 70$ años.
3. $\alpha = 0.05$.
4. Región crítica: $z > 1.645$, donde $z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}$.
5. Cálculos: $\bar{x} = 71.8$ años, $\sigma = 8.9$ años, en consecuencia, $z = \frac{71.8 - 70}{8.9/\sqrt{100}} = 2.02$.
6. Decisión: rechazar H_0 y concluir que la vida media actual es mayor que 70 años.

El valor P que corresponde a $z = 2.02$ es dado por el área de la región sombreada en la figura 10.10.

Si usamos la tabla A.3, tenemos

$$P = P(Z > 2.02) = 0.0217.$$

Como resultado, la evidencia a favor de H_1 es incluso más firme que la sugerida por un nivel de significancia de 0.05.

Ejemplo 10.4: Un fabricante de equipo deportivo desarrolló un nuevo sedal para pesca sintético que, según afirma, tiene una resistencia media a la rotura de 8 kilogramos con una desviación estándar de 0.5 kilogramos. Pruebe la hipótesis de que $\mu = 8$ kilogramos contra la alternativa de que $\mu \neq 8$ kilogramos si se prueba una muestra aleatoria de 50 sedales y se encuentra que tienen una resistencia media a la rotura de 7.8 kilogramos. Utilice un nivel de significancia de 0.01.

***Solución*:**
1. H_0: $\mu = 8$ kilogramos.
2. H_1: $\mu \neq 8$ kilogramos.
3. $\alpha = 0.01$.
4. Región crítica: $z < -2.575$ y $z > 2.575$, donde $z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}$.
5. Cálculos: $\bar{x} = 7.8$ kilogramos, $n = 50$, en consecuencia, $z = \frac{7.8 - 8}{0.5/\sqrt{50}} = -2.83$.
6. Decisión: rechazar H_0 y concluir que la resistencia promedio a la rotura no es igual a 8 sino que, de hecho, es menor que 8 kilogramos.

Como la prueba en este ejemplo es de dos colas, el valor de P que se desea es el doble del área de la región sombreada en la figura 10.11 a la izquierda de $z = -2.83$. Por lo tanto, si usamos la tabla A.3, tenemos

$$P = P(|Z| > 2.83) = 2P(Z < -2.83) = 0.0046,$$

que nos permite rechazar la hipótesis nula de que $\mu = 8$ kilogramos a un nivel de significancia menor que 0.01.

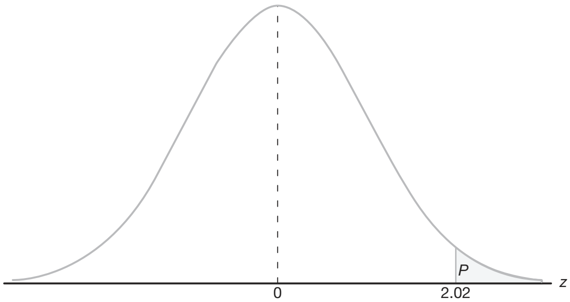

Figura 10.10: Valor P para el ejemplo 10.3.

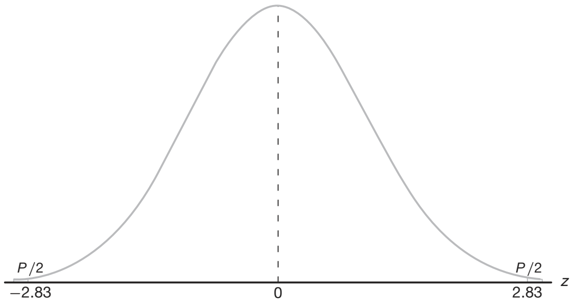

Figura 10.11: Valor P para el ejemplo 10.4.

Relación con la estimación del intervalo de confianza

El lector ya se habrá dado cuenta de que el método de la prueba de hipótesis para la inferencia estadística de este capítulo está muy relacionado con el método del intervalo de confianza del capítulo 9. La estimación del intervalo de confianza incluye el cálculo de límites dentro de los cuales es "razonable" que resida el parámetro en cuestión. Para el caso de una sola media de la población μ con σ^2 conocida, la estructura tanto de la prueba de hipótesis como de la estimación del intervalo de confianza se basa en la variable aleatoria

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}.$$

Resulta que la prueba de H_0: $\mu = \mu_0$ contra H_1: $\mu \neq \mu_0$ a un nivel de significancia α es *equivalente a calcular un intervalo de confianza del* $100(1-\alpha)\%$ *sobre* μ *y rechazar* H_0, *si* μ_0 *está fuera del intervalo de confianza.* Si μ_0 está dentro del intervalo de confianza, no se rechaza la hipótesis. La equivalencia es muy intuitiva y se puede ilustrar de manera muy simple. Recuerde que con un valor observado $\bar{x}$, no rechazar H_0 a un nivel de significancia α implica que

$$-z_{\alpha/2} \leq \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} \leq z_{\alpha/2},$$

que es equivalente a

$$\bar{x} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \leq \mu_0 \leq \bar{x} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}.$$

La equivalencia de la estimación del intervalo de confianza con la prueba de hipótesis se extiende a las diferencias entre dos medias, varianzas, cocientes de varianzas, etcétera. Como resultado, el estudiante de estadística no debería considerar la estimación del intervalo de confianza y la prueba de hipótesis como formas separadas de inferencia estadística. Considere el ejemplo 9.2 de la página 271. El intervalo de confianza del 95% sobre la media es dado por los límites (2.50, 2.70). Por consiguiente, con la misma información muestral, no se rechazará una hipótesis bilateral sobre μ que incluya cualquier valor hipotético entre 2.50 y 2.70. A medida que exploremos diferentes áreas de la prueba de hipótesis seguiremos aplicando la equivalencia a la estimación del intervalo de confianza.

Pruebas sobre una sola media (varianza desconocida)

Ciertamente sospecharíamos que las pruebas sobre una media de la población μ con σ^2 desconocida, como la estimación del intervalo de confianza, deberían incluir el uso de la distribución t de Student. En términos estrictos, la aplicación de la t de Student tanto para los intervalos de confianza como para la prueba de hipótesis se desarrolla bajo los siguientes supuestos. Las variables aleatorias $X_1, X_2, \ldots, X_n$ representan una muestra aleatoria de una distribución normal con μ y σ^2 desconocidas. Entonces, la variable aleatoria $\sqrt{n}(\bar{X}-\mu)/S$ tiene una distribución t de Student con $n-1$ grados de libertad. La estructura de la prueba es idéntica a la del caso en el que se conoce σ, excepto que el valor σ en el estadístico de prueba se reemplaza con el estimado calculado de S y la distribución normal estándar se reemplaza con una distribución t.

El estadístico t para una prueba sobre una sola media (varianza desconocida)

Para la hipótesis bilateral

$$H_0: \mu = \mu_0,$$

$$H_1: \mu \neq \mu_0,$$

rechazamos H_0 a un nivel de significancia α cuando el estadístico t calculado

$$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$$

excede a $t_{\alpha/2,n-1}$ o es menor que $-t_{\alpha/2,n-1}$.

El lector debería recordar de los capítulos 8 y 9 que la distribución t es simétrica alrededor del valor cero. Así, esta región crítica de dos colas se aplica de manera similar a la del caso en que se conoce σ. Para la hipótesis bilateral a un nivel de significancia α se aplican las regiones críticas de dos colas. Para H_1: $\mu > \mu_0$ el rechazo resulta cuando $t > t_{\alpha,n-1}$. Para H_1: $\mu < \mu_0$ la región crítica es dada por $T < -t_{\alpha,n-1}$.

Ejemplo 10.5: El Edison Electric Institute publica cifras del número de kilowatts-hora que gastan anualmente varios aparatos electrodomésticos. Se afirma que una aspiradora gasta un promedio de 46 kilowatts-hora al año. Si una muestra aleatoria de 12 hogares, que se incluye en un estudio planeado, indica que las aspiradoras gastan un promedio de 42 kilowatts-hora al año con una desviación estándar de 11.9 kilowatts-hora, ¿esto sugiere que las aspiradoras gastan, en promedio, menos de 46 kilowatts-hora al año a un nivel de significancia de 0.05? Suponga que la población de kilowatts-hora es normal.

***Solución*:**

1. H_0: $\mu = 46$ kilowatts-hora.
2. H_1: $\mu < 46$ kilowatts-hora.
3. $\alpha = 0.05$.
4. Región crítica: $t < -1.796$, donde $t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$ con 11 grados de libertad.
5. Cálculos: $\bar{x} = 42$ kilowatts-hora, $s = 11.9$ kilowatts-hora y $n = 12$.
   En consecuencia,

$$t = \frac{42 - 46}{11.9/\sqrt{12}} = -1.16, \qquad P = P(T < -1.16) \approx 0.135.$$

6. Decisión: no rechazar H_0 y concluir que el número promedio de kilowatts-hora que gastan al año las aspiradoras domésticas no es significativamente menor que 46. ▮

Comentario sobre la prueba *t* de una sola muestra

Es probable que el lector haya observado que se mantiene la equivalencia de la prueba t de dos colas para una sola media y el cálculo de un intervalo de confianza sobre μ con σ reemplazada por s. Considere el ejemplo 9.5 de la página 275. En esencia, podemos ver ese cálculo como uno en el que encontramos todos los valores de μ_0, el volumen medio hipotético de contenedores de ácido sulfúrico para los que la hipótesis H_0: $\mu = \mu_0$ *no se rechazará* con $\alpha = 0.05$. De nuevo, esto es consistente con el planteamiento: "si nos basamos en la información muestral, son razonables los valores del volumen medio de la población entre 9.74 y 10.26 litros".

En este punto vale la pena destacar algunos comentarios respecto a la suposición de normalidad. Indicamos que cuando se conoce σ, el teorema del límite central permite utilizar un estadístico de prueba o un intervalo de confianza que se base en Z, la variable aleatoria normal estándar. En términos estrictos, por supuesto, el teorema del límite central y, por lo tanto, el uso de la distribución normal estándar, no se aplica a menos que se conozca σ. En el capítulo 8 se estudió el desarrollo de la distribución t. Ahí se estableció que la normalidad sobre $X_1, X_2,..., X_n$ era una suposición subyacente. Entonces, *en sentido estricto,* no se deberían utilizar las tablas de t de Student de puntos porcentuales para pruebas o intervalos de confianza, a menos que se sepa que la muestra proviene de una población normal. En la práctica rara vez se puede suponer una σ conocida. Sin embargo, se dispondría de una buena estimación a partir de experimentos anteriores. Muchos libros de estadística sugieren que, cuando $n \geq 30$, es posible reemplazar con seguridad σ por s en el estadístico de prueba

$$z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}$$

con una población que tiene forma de campana y aun así utilizar las tablas Z para la región crítica adecuada. Aquí la implicación es que en realidad se recurre al teorema del límite central y que se confía en el hecho de que $s \approx \sigma$. Evidentemente, cuando se hace esto el resultado debe considerarse como una aproximación. Por consiguiente, un valor P calculado (de la distribución Z) de 0.15 puede ser 0.12 o quizá 0.17; o un intervalo de confianza calculado puede ser un intervalo de 93% de confianza en vez de un intervalo de 95% como se desea. Entonces, ¿qué sucede en las situaciones donde $n \leq 30$? El usuario no puede confiar en que s se acerque a σ, y para tomar en cuenta la inexactitud de la estimación el intervalo de confianza debería ser más ancho o el valor crítico de mayor magnitud. Los puntos porcentuales de la distribución t logran esto, pero sólo son correctos cuando la muestra proviene de una distribución normal. Desde luego, se pueden utilizar las gráficas de probabilidad normal para tener cierta idea de la desviación de la normalidad en un conjunto de datos.

Para muestras pequeñas a menudo resulta difícil detectar desviaciones de una distribución normal. (Las pruebas de la bondad del ajuste se presentan en una sección posterior de este capítulo). Para distribuciones en forma de campana de las variables aleatorias $X_1, X_2,..., X_n$, es probable que el uso de la distribución t para pruebas o intervalos de confianza produzca resultados muy buenos. Cuando haya duda, el usuario debería recurrir a los procedimientos no paramétricos que se presentan en el capítulo 16.

Impresiones o salidas por computadora con comentarios para pruebas *t* de una sola muestra

Seguramente al lector le interesará ver comentarios impresos por computadora que muestren el resultado de una prueba *t* con una sola muestra. Suponga que un ingeniero se interesa en probar el sesgo en un medidor de pH. Se reúnen datos de una sustancia neutra (pH = 7.0). Se toma una muestra de las mediciones y los datos son los siguientes:

7.07 7.00 7.10 6.97 7.00 7.03 7.01 7.01 6.98 7.08

Entonces, es de interés probar

$$H_0: \mu = 7.0,$$

$$H_1: \mu \neq 7.0.$$

En este caso utilizamos el paquete de cómputo *MINITAB* para ilustrar el análisis del conjunto de datos anterior. Observe los componentes clave de la impresión o salida que se muestra en la figura 10.12. Desde luego, la media $\bar{y}$ es 7.0250, StDev es simplemente la desviación estándar de la muestra $s = 0.044$ y SE Mean es el error estándar estimado de la media, y se calcula como $s/\sqrt{n} = 0.0139$. El valor *t* es el cociente

$$(7.0250-7)/0.0139 = 1.80.$$

```
pH-meter
  7.07   7.00   7.10   6.97   7.00   7.03   7.01   7.01   6.98   7.08
MTB > Onet 'pH-meter'; SUBC>  Test7.

One-Sample T: pH-meter Test of mu = 7 vs not = 7
Variable   N   Mean     StDev   SE Mean        95% CI          T     P
pH-meter  10 7.02500  0.04403  0.01392  (6.99350,7.05650) 1.80  0.106
```

Figura 10.12: Impresión de *MINITAB* para la prueba *t* de una muestra para el medidor de pH.

El valor *P* de 0.106 sugiere resultados que no son concluyentes. No hay evidencia que sugiera un firme rechazo de H_0 (con base en una α de 0.05 o de 0.10), **ni se puede concluir con certeza que el medidor de pH esté libre de sesgo**. Observe que el tamaño de la muestra de 10 es muy pequeño. Un incremento en el tamaño de la muestra (quizás otro experimento) podría resolver las cosas. En la sección 10.6 aparece un análisis respecto al tamaño adecuado de la muestra.

10.5 Dos muestras: pruebas sobre dos medias

El lector deberá comprender la relación entre pruebas e intervalos de confianza y sólo puede confiar plenamente en los detalles que ofrece el material sobre el intervalo de confianza del capítulo 9. Las pruebas respecto a dos medias representan un conjunto de herramientas analíticas muy importantes para el científico o el ingeniero. El procedimiento experimental es muy parecido al que se describe en la sección 9.8. Se extraen dos muestras aleatorias independientes de tamaños n_1 y n_2, respectivamente, de dos poblaciones con medias μ_1 y μ_2, y varianzas σ_1^2 y σ_2^2. Sabemos que la variable aleatoria

$$Z = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}}$$

tiene una distribución normal estándar. Suponemos aquí que n_1 y n_2 son suficientemente grandes, por lo que se aplica el teorema del límite central. Por supuesto, si las dos poblaciones son normales, el estadístico anterior tiene una distribución normal estándar incluso para n_1 y n_2 pequeñas. Evidentemente, si podemos suponer que $\sigma_1 = \sigma_2 = \sigma$, el estadístico anterior se reduce a

$$Z = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sigma\sqrt{1/n_1 + 1/n_2}}.$$

Los dos estadísticos anteriores sirven como base para el desarrollo de los procedimientos de prueba que incluyen dos medias. La equivalencia entre las pruebas y los intervalos de confianza, junto con los detalles técnicos implicados en las pruebas sobre una media, permiten que la transición a pruebas con dos medias sea sencilla.

La hipótesis bilateral sobre dos medias se escribe de manera muy general como

$$H_0\colon \mu_1 - \mu_2 = d_0.$$

Es evidente que la alternativa puede ser bilateral o unilateral. De nuevo, la distribución que se utiliza es la distribución del estadístico de prueba bajo H_0. Se calculan los valores $\bar{x}_1$ y $\bar{x}_2$, y para σ_1 y σ_2 conocidas, el estadístico de prueba es dado por

$$z = \frac{(\bar{x}_1 - \bar{x}_2) - d_0}{\sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}},$$

con una región crítica de dos colas en el caso de una alternativa bilateral. Es decir, se rechaza H_0 a favor de H_1: $\mu_1 - \mu_2 \neq d_0$, si $z > z_{\alpha/2}$ o $z < -z_{\alpha/2}$. Las regiones críticas de una cola se utilizan en el caso de alternativas unilaterales. El lector debería estudiar, como antes, el estadístico de prueba y estar satisfecho de que para, digamos H_1: $\mu_1 - \mu_2 > d_0$, la señal que favorece H_1 provenga de valores grandes de z. Por consiguiente, se aplica la región crítica de la cola superior.

Varianzas desconocidas pero iguales

Las situaciones más comunes que implican pruebas sobre dos medias son aquellas con varianzas desconocidas. Si el científico interesado está dispuesto a suponer que ambas distribuciones son normales y que $\sigma_1 = \sigma_2 = \sigma$, se puede utilizar la *prueba t agrupada* (a menudo llamada prueba *t* de dos muestras). El estadístico de prueba (véase la sección 9.8) es dado por el siguiente procedimiento de prueba.

Prueba t agrupada de dos muestras

Para la hipótesis bilateral

$$H_0: \mu_1 = \mu_2,$$
$$H_1: \mu_1 \neq \mu_2,$$

rechazamos H_0 al nivel de significancia α cuando el estadístico t calculado

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - d_0}{s_p\sqrt{1/n_1 + 1/n_2}},$$

donde

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

excede a $t_{\alpha/2,\, n_1+n_2-2}$ o es menor que $-t_{\alpha/2,\, n_1+n_2-2}$.

Recuerde que en el capítulo 9 se explicó que los grados de libertad para la distribución t son un resultado del agrupamiento de la información de las dos muestras para estimar σ^2. Las alternativas unilaterales, como era de esperarse, sugieren regiones críticas unilaterales. Por ejemplo, para H_1: $\mu_1 - \mu_2 > d_0$, rechace H_1: $\mu_1 - \mu_2 = d_0$ cuando $t > t_{\alpha,\, n_1+n_2-2}$.

Ejemplo 10.6: Se llevó a cabo un experimento para comparar el desgaste por abrasivos de dos diferentes materiales laminados. Se probaron 12 piezas del material 1 exponiendo cada pieza a una máquina para medir el desgaste. Se probaron 10 piezas del material 2 de manera similar. En cada caso se observó la profundidad del desgaste. Las muestras del material 1 revelaron un desgaste promedio (codificado) de 85 unidades con una desviación estándar muestral de 4; en tanto que las muestras del material 2 revelaron un promedio de 81 y una desviación estándar muestral de 5. ¿Podríamos concluir, a un nivel de significancia de 0.05, que el desgaste abrasivo del material 1 excede al del material 2 en más de 2 unidades? Suponga que las poblaciones son aproximadamente normales con varianzas iguales.

***Solución*:** Representemos con μ_1 y μ_2 las medias de la población del desgaste abrasivo para el material 1 y el material 2, respectivamente.

1. $H_0: \mu_1 - \mu_2 = 2$.
2. $H_1: \mu_1 - \mu_2 > 2$.
3. $\alpha = 0.05$.
4. Región crítica: $t > 1.725$, donde $t = \frac{(\bar{x}_1 - \bar{x}_2) - d_0}{s_p\sqrt{1/n_1 + 1/n_2}}$ con $v = 20$ grados de libertad.
5. Cálculos:

$$\bar{x}_1 = 85, \quad s_1 = 4, \quad n_1 = 12,$$
$$\bar{x}_2 = 81, \quad s_2 = 5, \quad n_2 = 10.$$

En consecuencia,

$$s_p = \sqrt{\frac{(11)(16) + (9)(25)}{12 + 10 - 2}} = 4.478,$$

$$t = \frac{(85 - 81) - 2}{4.478\sqrt{1/12 + 1/10}} = 1.04,$$

$$P = P(T > 1.04) \approx 0.16. \qquad \text{(Véase la tabla A.4).}$$

6. Decisión: no rechazar H_0. No podemos concluir que el desgaste abrasivo del material 1 excede al del material 2 en más de 2 unidades.

Varianzas desconocidas pero diferentes

Hay situaciones donde al analista **no** le es posible suponer que $\sigma_1 = \sigma_2$. De la sección 9.8 recuerde que, si las poblaciones son normales, el estadístico

$$T' = \frac{(\bar{X}_1 - \bar{X}_2) - d_0}{\sqrt{s_1^2/n_1 + s_2^2/n_2}}$$

tiene una distribución *t* aproximada con grados de libertad aproximados

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

Como resultado, el procedimiento de prueba consiste en *no rechazar* H_0 cuando

$$-t_{\alpha/2,v} < t' < t_{\alpha/2,v},$$

con v dado como antes. De nuevo, como en el caso de la prueba *t* agrupada, las alternativas unilaterales sugieren regiones críticas unilaterales.

Observaciones pareadas

Un estudio de la prueba *t* de dos muestras o el intervalo de confianza sobre la diferencia entre medias deberían sugerir la necesidad de un diseño experimental. Recuerde el análisis de las unidades experimentales en el capítulo 9, donde se sugirió que las condiciones de las dos poblaciones (a menudo denominadas como los dos tratamientos) se deberían asignar de manera aleatoria a las unidades experimentales. Esto se realiza para evitar resultados sesgados debido a las diferencias sistemáticas entre unidades experimentales. En otras palabras, en términos de la jerga para la prueba de hipótesis, es importante que la diferencia significativa que se encuentre entre las medias se deba a las diferentes condiciones de las poblaciones y no a las unidades experimentales en el estudio. Por ejemplo, considere el ejercicio 9.40 de la sección 9.9. Los 20 tallos desempeñan el papel de unidades experimentales. Diez de ellos se tratan con nitrógeno y 10 se dejan sin tratamiento. Es muy importante que esta asignación a los tratamientos "con nitrógeno" y "sin nitrógeno" sea aleatoria para garantizar que las diferencias sistemáticas entre los tallos no interfieran con una comparación válida entre las medias.

En el ejemplo 10.6 el momento de la medición es la opción más probable de la unidad experimental. Las 22 piezas de material se deberían medir en orden aleatorio.

Necesitamos protegernos contra la posibilidad de que las mediciones del desgaste que se realicen casi al mismo tiempo tiendan a dar resultados similares. *No se esperan* **diferencias sistemáticas** (no aleatorias) **en las unidades experimentales**. Sin embargo, las asignaciones aleatorias protegen contra el problema.

Las referencias a la planeación de experimentos, aleatorización, elección del tamaño de la muestra, etcétera, continuarán influyendo en gran parte del desarrollo en los capítulos 13, 14 y 15. Cualquier científico o ingeniero cuyo interés resida en el análisis de datos reales debería estudiar este material. La prueba t agrupada se amplía en el capítulo 13 para cubrir más de dos medias.

La prueba de dos medias se puede llevar a cabo cuando los datos están en forma de observaciones pareadas, como se estudió en el capítulo 9. En esta estructura de pareado las condiciones de las dos poblaciones (tratamientos) se asignan de forma aleatoria dentro de unidades homogéneas. El cálculo del intervalo de confianza para $\mu_1 - \mu_2$ en la situación con observaciones pareadas se basa en la variable aleatoria

$$T = \frac{\bar{D} - \mu_D}{S_d/\sqrt{n}},$$

donde $\bar{D}$ y S_d son variables aleatorias que representan la media muestral y la desviación estándar de las diferencias de las observaciones en las unidades experimentales. Como en el caso de la *prueba t agrupada,* la suposición es que las observaciones de cada población son normales. Este problema de dos muestras se reduce en esencia a un problema de una muestra utilizando las diferencias calculadas $d_1, d_2, \ldots, d_n$. Por consiguiente, la hipótesis se reduce a

$$H_0\colon\ \mu_D = d_0.$$

El estadístico de prueba calculado es dado entonces por

$$t = \frac{\bar{d} - d_0}{s_d/\sqrt{n}}.$$

Las regiones críticas se construyen usando la distribución t con $n - 1$ grados de libertad.

El problema de la interacción en una prueba *t* pareada

El siguiente estudio de caso no sólo ilustra el uso de la prueba t pareada, sino que el análisis revelará mucho sobre las dificultades que surgen cuando ocurre una interacción entre los tratamientos y las unidades experimentales en la estructura de la t pareada. Recuerde que la interacción entre factores se presentó en la sección 1.7, en un análisis de los tipos generales de estudios estadísticos. El concepto de interacción será un tema importante desde el capítulo 13 hasta el 15.

Existen ciertos tipos de pruebas estadísticas en los que la existencia de una interacción produce dificultades. Un ejemplo es la prueba t pareada. En la sección 9.9 se utilizó la estructura pareada en el cálculo de un intervalo de confianza sobre la diferencia entre dos medias, y se reveló la ventaja del pareado para situaciones en que las unidades experimentales son homogéneas. El pareado produce una reducción en σ_D, la desviación estándar de una diferencia $D_i = X_{1i} - X_{2i}$, como se explicó en la sección 9.9. Si hay una interacción entre los tratamientos y las unidades experimentales, la ventaja lograda mediante el pareado se podría reducir de manera sustancial. Por consiguiente, en el ejemplo 9.13 de la página 293 la suposición de la ausencia de interacción permitió que la diferencia en los niveles medios de TCDD (plasma contra tejido adiposo) fuera la misma en todos los veteranos. Un vistazo rápido a los datos sugiere que no hay una violación significativa de los supuestos de ausencia de interacción.

Para demostrar cómo influye la interacción en Var(D) y, por lo tanto, en la calidad de la prueba t pareada, es aleccionador revisar la i-ésima diferencia dada por $D_i = X_{1i} - X_{2i} = (\mu_1 - \mu_2) + (\epsilon_1 - \epsilon_2)$, donde X_{1i} y X_{2i} se toman de la i-ésima unidad experimental. Si la unidad pareada es homogénea, los errores en X_{1i} y en X_{2i} serán similares y no independientes. En el capítulo 9 señalamos que la covarianza positiva entre los errores da como resultado una Var(D) reducida. Por consiguiente, el tamaño de la diferencia en los tratamientos y la relación entre los errores en X_{1i} y X_{2i}, a los que contribuye la unidad experimental, tenderán a permitir la detección de una diferencia significativa.

¿Qué condiciones resultan en una interacción?

Consideremos una situación en la que las unidades experimentales no son homogéneas. Más bien, considere la i-ésima unidad experimental con las variables aleatorias X_{1i} y X_{2i} que no son similares. Sean ϵ_{1i} y ϵ_{2i} variables aleatorias que representan los errores en los valores X_{1i} y X_{2i}, respectivamente, en la unidad i-ésima. Así, podemos escribir

$$X_{1i} = \mu_1 + \epsilon_{1i} \text{ y } X_{2i} = \mu_2 + \epsilon_{2i}.$$

Los errores con valor esperado cero podrían tender a provocar que los valores de respuesta X_{1i} y X_{2i} se muevan en direcciones opuestas, dando como resultado un valor negativo para Cov(ϵ_{1i}, ϵ_{2i}) y, por ende, un valor negativo para Cov(X_{1i}, X_{2i}). En realidad, el modelo se podría volver aún más complicado por el hecho de que $\sigma_1^2 = \text{Var}(\epsilon_{1i}) \neq \sigma_2^2 = \text{Var}(\epsilon_{2i})$. Los parámetros de la varianza y la covarianza podrían variar entre las n unidades experimentales. Así, a diferencia del caso con homogeneidad, D_i tenderá a ser muy diferente en todas las unidades experimentales debido a la naturaleza heterogénea de la diferencia en $\epsilon_1 - \epsilon_2$ entre las unidades. Esto produce la interacción entre los tratamientos y las unidades. Además, para una unidad experimental específica (véase el teorema 4.9),

$$\sigma_D^2 = \text{Var}(D) = \text{Var}(\epsilon_1) + \text{Var}(\epsilon_2) - 2\,\text{Cov}(\epsilon_1, \epsilon_2)$$

está inflado por el término negativo de covarianza, de manera que la ventaja lograda por el pareado en el caso de la unidad homogénea se pierde en el caso que aquí se describe. En tanto que la inflación en Var(D) variará de un caso a otro, en algunas situaciones existe el peligro de que el aumento en la varianza neutralice cualquier diferencia que exista entre μ_1 y μ_2. Desde luego, un valor grande de $\bar{d}$ en el estadístico t podría reflejar una diferencia en el tratamiento que compense el estimado inflado de la varianza s_d^2.

Estudio de caso 10.1: **Datos de muestra de sangre:** En un estudio realizado en el Departamento de Silvicultura y Fauna de Virginia Tech, J. A. Wesson examinó la influencia del fármaco *succinylcholine* sobre los niveles de circulación de andrógenos en la sangre. Se obtuvieron muestras de sangre de venados salvajes inmediatamente después de recibir una inyección intramuscular de *succinylcholine* con dardos de un rifle de caza. Treinta minutos después se obtuvo una segunda muestra de sangre y después los venados fueron liberados. Los niveles de andrógenos de 15 venados al momento de la captura y 30 minutos más tarde, medidos en nanogramos por mililitro (ng/mL), se presentan en la tabla 10.2.

Suponga que las poblaciones de niveles de andrógenos al momento de la inyección y 30 minutos después se distribuyen normalmente, y pruebe, a un nivel de significancia de 0.05, si las concentraciones de andrógenos se alteraron después de 30 minutos.

Tabla 10.2: Datos para el estudio de caso 10.1

	Andrógenos (ng/mL)		
Venado	**Al momento de la inyección**	**30 minutos después de la inyección**	d_i
1	2.76	7.02	4.26
2	5.18	3.10	−2.08
3	2.68	5.44	2.76
4	3.05	3.99	0.94
5	4.10	5.21	1.11
6	7.05	10.26	3.21
7	6.60	13.91	7.31
8	4.79	18.53	13.74
9	7.39	7.91	0.52
10	7.30	4.85	−2.45
11	11.78	11.10	−0.68
12	3.90	3.74	−0.16
13	26.00	94.03	68.03
14	67.48	94.03	26.55
15	17.04	41.70	24.66

***Solución*:** Sean μ_1 y μ_2 la concentración promedio de andrógenos al momento de la inyección y 30 minutos después, respectivamente. Procedemos como sigue:

1. H_0: $\mu_1 = \mu_2$ o $\mu_D = \mu_1 - \mu_2 = 0$.

2. H_1: $\mu_1 \neq \mu_2$ o $\mu_D = \mu_1 - \mu_2 \neq 0$.

3. $\alpha = 0.05$.

4. Región crítica: $t < -2.145$ y $t > 2.145$, donde $t = \frac{\bar{d} - d_0}{s_D/\sqrt{n}}$ con $v = 14$ grados de libertad.

5. Cálculos: La media muestral y la desviación estándar para las d_i son

$$\bar{d} = 9.848 \text{ y } s_d = 18.474.$$

Por lo tanto,

$$t = \frac{9.848 - 0}{18.474/\sqrt{15}} = 2.06.$$

6. Aunque el estadístico t no es significativo al nivel 0.05, de la tabla A.4,

$$P = P(|T| > 2.06) \approx 0.06.$$

Como resultado, existe cierta evidencia de que hay una diferencia en los niveles medios circulantes de andrógenos. ▟

La suposición de la ausencia de interacción implicaría que el efecto sobre los niveles de andrógenos de los venados es casi el mismo en los datos de ambos tratamientos, es decir, en el momento de la inyección de *succinylcholine* y 30 minutos después. Esto se puede expresar cambiando los papeles de los dos factores; por ejemplo, la diferencia en los tratamientos es casi igual en todas las unidades, es decir, los venados. Ciertamente hay algunas combinaciones venado/tratamiento para las que parece ser válida la suposición de ausencia de interacción, pero difícilmente existen evidencias firmes de que las unidades experimentales sean homogéneas. Sin embargo, la naturaleza de la interacción y el incremento resultante en Var($\bar{D}$) parecen estar dominados por una diferencia sustancial en los tratamientos. Esto también es demostrado por el hecho de que 11 de los 15 venados mostraron señales positivas para las d_i calculadas y las d_i negativas (para los venados 2, 10, 11 y 12) son pequeñas en magnitud comparadas con las 12 positivas. Por consiguiente, al parecer el nivel medio de andrógenos es significativamente más alto 30 minutos después de la inyección que en el momento en que se aplica, y las conclusiones podrían ser más firmes de lo que sugiere $p = 0.06$.

Impresiones por computadora con comentarios para pruebas *t* pareadas

La figura 10.13 presenta una impresión por computadora del *SAS* para una prueba *t* pareada usando los datos del estudio de caso 10.1. Observe que el listado se parece al de una prueba *t* de una sola muestra y, por supuesto, esto es con exactitud lo que se realizó, ya que la prueba busca determinar si $\bar{d}$ es significativamente diferente de cero.

```
                Analysis Variable : Diff

  N           Mean      Std Error   t Value   Pr > |t|
 -------------------------------------------------------
 15      9.8480000     4.7698699     2.06     0.0580
 -------------------------------------------------------
```

Figura 10.13: Impresión por computadora del *SAS* de la prueba *t* pareada para los datos del estudio de caso 10.1.

Resumen de los procedimientos de prueba

Mientras completamos el desarrollo formal de pruebas sobre medias de la población, ofrecemos la tabla 10.3, que resume el procedimiento de prueba para los casos de una sola media y de dos medias. Observe el procedimiento aproximado cuando las distribuciones son normales y las varianzas se desconocen pero no se suponen iguales. Este estadístico se estudió en el capítulo 9.

10.6 Elección del tamaño de la muestra para la prueba de medias

En la sección 10.2 demostramos cómo el analista puede explotar las relaciones entre el tamaño de la muestra, el nivel de significancia α y la potencia de la prueba para alcanzar cierto estándar de calidad. En la mayoría de las circunstancias prácticas el experimento debería planearse y, de ser posible, elegir el tamaño de la muestra antes del proceso de recolección de datos. Por lo general el tamaño de la muestra se determina de modo que

Tabla 10.3: Pruebas relacionadas con medias

H_0	Valor del estadístico de prueba	H_1	Región crítica
$\mu = \mu_0$	$z = \dfrac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}$; σ conocida	$\mu < \mu_0$ $\mu > \mu_0$ $\mu \neq \mu_0$	$z < -z_\alpha$ $z > z_\alpha$ $z < -z_{\alpha/2}$ o $z > z_{\alpha/2}$
$\mu = \mu_0$	$t = \dfrac{\bar{x} - \mu_0}{s/\sqrt{n}}$; $v = n - 1$, σ desconocida	$\mu < \mu_0$ $\mu > \mu_0$ $\mu \neq \mu_0$	$t < -t_\alpha$ $t > t_\alpha$ $t < -t_{\alpha/2}$ o $t > t_{\alpha/2}$
$\mu_1 - \mu_2 = d_0$	$z = \dfrac{(\bar{x}_1 - \bar{x}_2) - d_0}{\sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}}$; σ_1 y σ_2 conocidas	$\mu_1 - \mu_2 < d_0$ $\mu_1 - \mu_2 > d_0$ $\mu_1 - \mu_2 \neq d_0$	$z < -z_\alpha$ $z > z_\alpha$ $z < -z_{\alpha/2}$ o $z > z_{\alpha/2}$
$\mu_1 - \mu_2 = d_0$	$t = \dfrac{(\bar{x}_1 - \bar{x}_2) - d_0}{s_p\sqrt{1/n_1 + 1/n_2}}$; $v = n_1 + n_2 - 2$, $\sigma_1 = \sigma_2$ pero desconocidas $s_p^2 = \dfrac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$	$\mu_1 - \mu_2 < d_0$ $\mu_1 - \mu_2 > d_0$ $\mu_1 - \mu_2 \neq d_0$	$t < -t_\alpha$ $t > t_\alpha$ $t < -t_{\alpha/2}$ o $t > t_{\alpha/2}$
$\mu_1 - \mu_2 = d_0$	$t' = \dfrac{(\bar{x}_1 - \bar{x}_2) - d_0}{\sqrt{s_1^2/n_1 + s_2^2/n_2}}$; $v = \dfrac{(s_1^2/n_1 + s_2^2/n_2)^2}{\frac{(s_1^2/n_1)^2}{n_1 - 1} + \frac{(s_2^2/n_2)^2}{n_2 - 1}}$, $\sigma_1 \neq \sigma_2$ y desconocidas	$\mu_1 - \mu_2 < d_0$ $\mu_1 - \mu_2 > d_0$ $\mu_1 - \mu_2 \neq d_0$	$t' < -t_\alpha$ $t' > t_\alpha$ $t' < -t_{\alpha/2}$ o $t' > t_{\alpha/2}$
$\mu_D = d_0$ observaciones pareadas	$t = \dfrac{\bar{d} - d_0}{s_d/\sqrt{n}}$; $v = n - 1$	$\mu_D < d_0$ $\mu_D > d_0$ $\mu_D \neq d_0$	$t < -t_\alpha$ $t > t_\alpha$ $t < -t_{\alpha/2}$ o $t > t_{\alpha/2}$

permita lograr una buena potencia para una α fija y una alternativa específica fija. Esta alternativa fija puede estar en la forma de $\mu - \mu_0$ en el caso de una hipótesis que incluya una sola media o $\mu_1 - \mu_2$ en el caso de un problema que implique dos medias. Los casos específicos serán ilustrativos.

Suponga que deseamos probar la hipótesis

$$H_0: \mu = \mu_0,$$

$$H_1 = \mu > \mu_0,$$

con un nivel de significancia α, cuando se conoce la varianza σ^2. Para una alternativa específica, digamos, $\mu = \mu_0 + \delta$, en la figura 10.14 se muestra que la potencia de nuestra prueba es

$$1 - \beta = P(\bar{X} > a \text{ cuando } \mu = \mu_0 + \delta).$$

Por lo tanto,

$$\begin{aligned}\beta &= P(\bar{X} < a \text{ cuando } \mu = \mu_0 + \delta)\\ &= P\left[\frac{\bar{X} - (\mu_0 + \delta)}{\sigma/\sqrt{n}} < \frac{a - (\mu_0 + \delta)}{\sigma/\sqrt{n}} \text{ cuando } \mu = \mu_0 + \delta\right].\end{aligned}$$

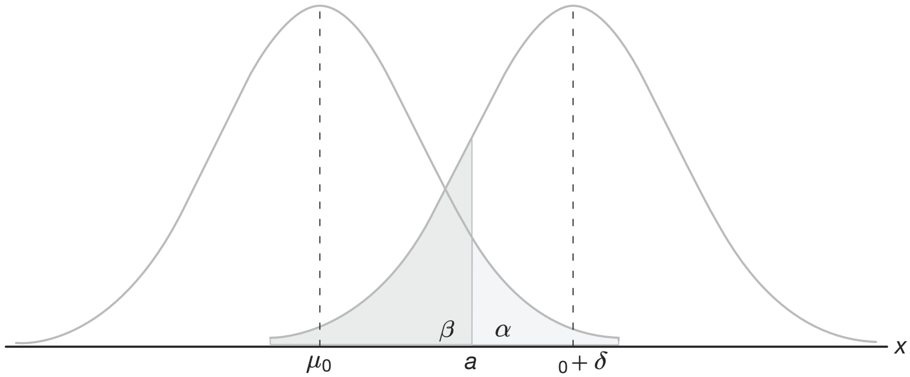

Figura 10.14: Prueba de $\mu = \mu_0$ contra $\mu = \mu_0 + \delta$.

Bajo la hipótesis alternativa $\mu = \mu_0 + \delta$, el estadístico

$$\frac{\bar{X} - (\mu_0 + \delta)}{\sigma/\sqrt{n}}$$

es la variable normal estándar Z. Por lo tanto,

$$\beta = P\left(Z < \frac{a - \mu_0}{\sigma/\sqrt{n}} - \frac{\delta}{\sigma/\sqrt{n}}\right) = P\left(Z < z_\alpha - \frac{\delta}{\sigma/\sqrt{n}}\right),$$

de donde concluimos que

$$-z_\beta = z_\alpha - \frac{\delta\sqrt{n}}{\sigma},$$

y, en consecuencia,

$$\text{Elección del tamaño de la muestra: } n = \frac{(z_\alpha + z_\beta)^2\sigma^2}{\delta^2},$$

un resultado que también es verdadero cuando la hipótesis alternativa es $\mu < \mu_0$.

En el caso de una prueba de dos colas obtenemos la potencia $1 - \beta$ para una alternativa específica cuando

$$n \approx \frac{(z_{\alpha/2} + z_\beta)^2\sigma^2}{\delta^2}.$$

Ejemplo 10.7: Suponga que deseamos probar la hipótesis

$$H_0\colon \mu = 68 \text{ kilogramos},$$
$$H_1\colon \mu > 68 \text{ kilogramos},$$

para los pesos de estudiantes hombres en cierta universidad usando un nivel de significancia $\alpha = 0.05$ cuando se sabe que $\sigma = 5$. Calcule el tamaño muestral que se requiere si la potencia de nuestra prueba debe ser 0.95 cuando la media real es 69 kilogramos.

Solución: Como $\alpha = \beta = 0.05$, tenemos $z_\alpha = z_\beta = 1.645$. Para la alternativa $\beta = 69$ tomamos $\delta = 1$ y entonces,

$$n = \frac{(1.645 + 1.645)^2(25)}{1} = 270.6.$$

Por lo tanto, se requieren 271 observaciones si la prueba debe rechazar la hipótesis nula el 95% de las veces cuando, de hecho, μ es tan grande como 69 kilogramos.

El caso de dos muestras

Se puede utilizar un procedimiento similar para determinar el tamaño de la muestra $n = n_1 = n_2$ que se requiere para una potencia específica de la prueba en que se comparan dos medias de la población. Por ejemplo, suponga que deseamos probar la hipótesis

$$H_0\colon \mu_1 - \mu_2 = d_0,$$
$$H_1\colon \mu_1 - \mu_2 \neq d_0,$$

cuando se conocen σ_1 y σ_2. Para una alternativa específica, digamos $\mu_1 - \mu_2 = d_0 + \delta$, en la figura 10.15 se muestra que la potencia de nuestra prueba es

$$1 - \beta = P(|\bar{X}_1 - \bar{X}_2| > a \text{ cuando } \mu_1 - \mu_2 = d_0 + \delta).$$

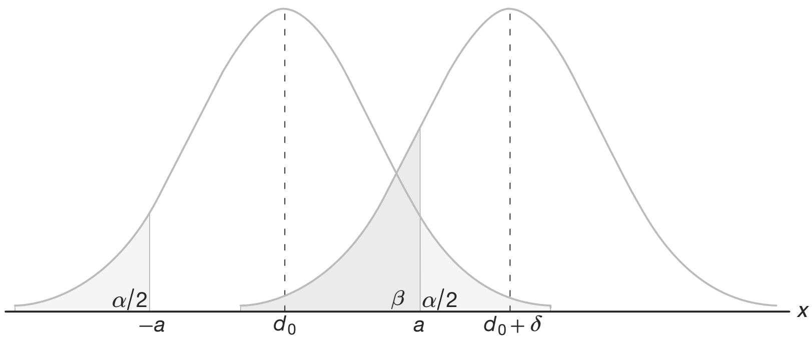

Figura 10.15: Prueba de $\mu_1 - \mu_2 = d_0$ contra $\mu_1 - \mu_2 = d_0 + \delta$.

Por lo tanto,

$$\begin{aligned}\beta &= P(-a < \bar{X}_1 - \bar{X}_2 < a \text{ cuando } \mu_1 - \mu_2 = d_0 + \delta)\\ &= P\left[\frac{-a - (d_0 + \delta)}{\sqrt{(\sigma_1^2 + \sigma_2^2)/n}} < \frac{(\bar{X}_1 - \bar{X}_2) - (d_0 + \delta)}{\sqrt{(\sigma_1^2 + \sigma_2^2)/n}}\right.\\ &\qquad \left. < \frac{a - (d_0 + \delta)}{\sqrt{(\sigma_1^2 + \sigma_2^2)/n}} \text{ cuando } \mu_1 - \mu_2 = d_0 + \delta\right].\end{aligned}$$

Con la hipótesis alternativa $\mu_1 - \mu_2 = d_0 + \delta$, el estadístico

$$\frac{\bar{X}_1 - \bar{X}_2 - (d_0 + \delta)}{\sqrt{(\sigma_1^2 + \sigma_2^2)/n}}$$

es la variable normal estándar Z. Ahora bien, al escribir

$$-z_{\alpha/2} = \frac{-a - d_0}{\sqrt{(\sigma_1^2 + \sigma_2^2)/n}} \quad \text{y} \quad z_{\alpha/2} = \frac{a - d_0}{\sqrt{(\sigma_1^2 + \sigma_2^2)/n}},$$

tenemos

$$\beta = P\left[-z_{\alpha/2} - \frac{\delta}{\sqrt{(\sigma_1^2 + \sigma_2^2)/n}} < Z < z_{\alpha/2} - \frac{\delta}{\sqrt{(\sigma_1^2 + \sigma_2^2)/n}}\right],$$

de donde concluimos que

$$-z_\beta \approx z_{\alpha/2} - \frac{\delta}{\sqrt{(\sigma_1^2 + \sigma_2^2)/n}},$$

y, por lo tanto,

$$n \approx \frac{(z_{\alpha/2} + z_\beta)^2(\sigma_1^2 + \sigma_2^2)}{\delta^2}.$$

Para la prueba de una sola cola, la expresión para el tamaño requerido de la muestra cuando $n = n_1 = n_2$ es

$$\text{Elección del tamaño de la muestra: } n = \frac{(z_\alpha + z_\beta)^2(\sigma_1^2 + \sigma_2^2)}{\delta^2}.$$

Cuando se desconoce la varianza de la población (o varianzas en la situación de dos muestras), la elección del tamaño de la muestra no es directa. Al probar la hipótesis $\mu = \mu_0$ cuando el valor verdadero es $\mu = \mu_0 + \delta$, el estadístico

$$\frac{\bar{X} - (\mu_0 + \delta)}{S/\sqrt{n}}$$

no sigue la distribución t, como se podría esperar, más bien sigue la **distribución *t* no central**. Sin embargo, existen tablas o gráficas que se basan en la distribución t no central para determinar el tamaño adecuado de la muestra, si se dispone de algún estimado de σ o si δ es un múltiplo de σ. La tabla A.8 proporciona los tamaños muestrales necesarios para controlar los valores de α y β para diversos valores de

$$\Delta = \frac{|\delta|}{\sigma} = \frac{|\mu - \mu_0|}{\sigma}$$

en el caso de pruebas de una y de dos colas. En el caso de la prueba t de dos muestras en la que se desconocen las varianzas pero se suponen iguales, obtenemos los tamaños muestrales $n = n_1 = n_2$ necesarios para controlar los valores de α y β para diversos valores de

$$\Delta = \frac{|\delta|}{\sigma} = \frac{|\mu_1 - \mu_2 - d_0|}{\sigma}$$

de la tabla A.9.

Ejemplo 10.8: Al comparar el comportamiento de dos catalizadores sobre el efecto del producto de una reacción se realiza una prueba t de dos muestras con $\alpha = 0.05$. Se considera que las varianzas de los productos son iguales para los dos catalizadores. ¿De qué tamaño debe ser una muestra para cada catalizador si se desea probar la hipótesis

$$H_0\colon \mu_1 = \mu_2,$$

$$H_1\colon \mu_1 \neq \mu_2$$

si es esencial detectar una diferencia de 0.8σ entre los catalizadores con 0.9 de probabilidad?

Solución: De la tabla A.9, con $\alpha = 0.05$ para una prueba de dos colas, $\beta = 0.1$ y

$$\Delta = \frac{|0.8\sigma|}{\sigma} = 0.8,$$

encontramos que el tamaño requerido de la muestra es $n = 34$.

En situaciones prácticas sería difícil forzar a un científico o a un ingeniero a hacer un compromiso sobre la información a partir de la cual se puede encontrar un valor de Δ. Se recuerda al lector que el valor Δ cuantifica el tipo de diferencia entre las medias que el científico considera importantes; es decir, una diferencia que se considere *significativa* desde un punto de vista científico, no estadístico. El ejemplo 10.8 ilustra cómo suele hacerse esta elección, a saber, mediante la selección de una fracción de σ. Evidentemente, si el tamaño de la muestra se basa en una elección de $|\delta|$, que es una fracción pequeña de σ, el tamaño muestral que resulta podría ser muy grande comparado con lo que permite el estudio.

10.7 Métodos gráficos para comparar medias

En el capítulo 1 se puso mucha atención a la presentación de datos en forma gráfica, como los diagramas de tallo y hojas y las gráficas de caja y bigote. En la sección 8.8 las gráficas de cuantiles y las gráficas normales cuantil-cuantil se utilizaron para brindar una "imagen" y resumir así un conjunto de datos experimentales. Muchos paquetes de cómputo producen representaciones gráficas. A medida que procedamos con otras formas de análisis de datos, por ejemplo, el análisis de regresión y el análisis de varianza, los métodos gráficos se vuelven aún más informativos.

Los auxiliares gráficos no se pueden utilizar como un reemplazo del propio procedimiento de prueba. En realidad, el valor del estadístico de prueba indica el tipo adecuado de evidencia en apoyo de H_0 o H_1. Sin embargo, una imagen ofrece una buena ilustración y a menudo es un mejor comunicador de evidencia para el beneficiario del análisis. Además, una imagen con frecuencia dejará claro por qué se encontró una diferencia significativa. La falla de una suposición importante se puede expresar mediante un resumen gráfico.

Para la comparación de medias, las gráficas de caja y bigote simultáneas proporcionan una imagen clara. El lector debería recordar que estas gráficas muestran el percentil 25, el percentil 75 y la mediana en un conjunto de datos. Además, las extensiones muestran los extremos en un conjunto de datos. Considere el ejercicio 10.40 al final de esta sección. Se midieron los niveles en plasma de ácido ascórbico en dos grupos de mujeres embarazadas: fumadoras y no fumadoras. En la figura 10.16 se observan las gráficas de caja y bigote para ambos grupos de mujeres y dos cosas son muy evidentes; al tomar en cuenta la variabilidad parece haber una diferencia despreciable en las medias muestrales. Además, parece que la variabilidad en los dos grupos es hasta cierto punto diferente. Desde luego, el analista debe tener en la mente las más bien considerables diferencias entre los tamaños muestrales en este caso.

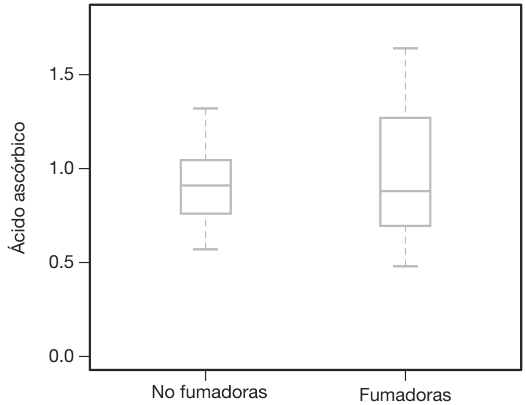

Figura 10.16: Dos gráficas de caja y bigote con los datos de ácido ascórbico para mujeres fumadoras y no fumadoras.

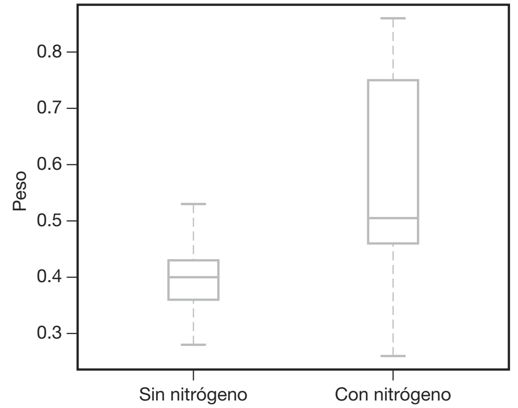

Figura 10.17: Dos gráficas de caja y bigote para los datos de los tallos.

Considere el ejercicio 9.40 de la sección 9.9. En la figura 10.17 se presenta la gráfica múltiple de caja y bigote para los datos de 10 tallos, de los cuales sólo la mitad recibió el tratamiento con nitrógeno. Tal gráfica revela una variabilidad menor para el grupo que no recibió nitrógeno. Además, la falta de traslape de las cajas sugiere una diferencia significativa entre los pesos medios de los tallos para los dos grupos. Parecería que la presencia de nitrógeno aumenta el peso de los tallos y quizás aumente la variabilidad en los pesos.

No existen reglas generales relacionadas con el momento cuando dos gráficas de caja y bigote brindan evidencia de diferencias significativas entre las medias. Sin embargo, una pauta aproximada es que si la línea del percentil 25 para una muestra excede a la línea de la mediana de la otra muestra, hay evidencia sólida de una diferencia entre las medias.

Se hará más énfasis en los métodos gráficos en un estudio de caso de la vida real que se presenta más adelante en este capítulo.

Impresiones por computadora con comentarios para pruebas *t* con dos muestras

Considere nuevamente el ejercicio 9.40 de la página 294, donde se reunieron datos de tallos que recibieron y no recibieron nitrógeno. Pruebe

$$H_0\colon \mu_{\text{NIT}} = \mu_{\text{NO}},$$

$$H_1\colon \mu_{\text{NIT}} > \mu_{\text{NO}},$$

donde las medias de la población indican los pesos medios. La figura 10.18 es una impresión por computadora con comentarios generados con el programa *SAS*. Observe que se presentan la desviación estándar y el error estándar muestrales para ambas muestras. También se incluye el estadístico t bajo la suposición de varianzas iguales y varianzas diferentes. En la gráfica de caja y bigote que se observa en la figura 10.17 en realidad parece que se transgrede la suposición de igualdad de varianzas. Un valor P de 0.0229 sugiere una conclusión de medias diferentes. Esto coincide con la información de diagnóstico que se presenta en la figura 10.18. A propósito, observe que t y t' son iguales en este caso, ya que $n_1 = n_2$.

```
                TTEST Procedure
Variable  Weight
    Mineral    N       Mean  Std Dev  Std Err
No nitrogen   10     0.3990   0.0728  0.0230
   Nitrogen   10     0.5650   0.1867  0.0591

Variances     DF    t Value   Pr >|t|
    Equal     18       2.62    0.0174
  Unequal   11.7       2.62    0.0229

          Test the Equality of Variances
 Variable  Num DF     Den DF    F Value  Pr > F
   Weight      9          9       6.58  0.0098
```

Figura 10.18: Impresión del *SAS* para la prueba *t* de dos muestras.

Ejercicios

10.19 En un informe de investigación, Richard H. Weindruch, de la Escuela de Medicina de la UCLA, afirma que los ratones con una vida promedio de 32 meses vivirán hasta alrededor de 40 meses si 40% de las calorías en su dieta se reemplazan con vitaminas y proteínas. ¿Hay alguna razón para creer que $\mu < 40$, si 64 ratones que son sometidos a esa dieta tienen una vida promedio de 38 meses, con una desviación estándar de 5.8 meses? Utilice un valor *P* en su conclusión.

10.20 Una muestra aleatoria de 64 bolsas de palomitas con queso chedar pesan, en promedio, 5.23 onzas, con una desviación estándar de 0.24 onzas. Pruebe la hipótesis de que $\mu = 5.5$ onzas contra la hipótesis alternativa de que $\mu < 5.5$ onzas, al nivel de significancia de 0.05.

10.21 Una empresa de material eléctrico fabrica bombillas que tienen una duración que se distribuye de forma aproximadamente normal con una media de 800 horas y una desviación estándar de 40 horas. Pruebe la hipótesis de que $\mu = 800$ horas contra la alternativa de que $\mu \neq 800$ horas, si una muestra aleatoria de 30 bombillas tiene una duración promedio de 788 horas. Utilice un valor *P* en su respuesta.

10.22 En la revista *Hypertension* de la American Heart Association, investigadores reportan que los individuos que practican la meditación trascendental (MT) bajan su presión sanguínea de forma significativa. Si una muestra aleatoria de 225 hombres que practican la MT meditan 8.5 horas a la semana, con una desviación estándar de 2.25 horas, ¿esto sugiere que, en promedio, los hombres que utilizan la MT meditan más de 8 horas por semana? Cite un valor *P* en su conclusión.

10.23 Pruebe la hipótesis de que el contenido promedio de los envases de un lubricante específico es de 10 litros, si los contenidos de una muestra aleatoria de 10 envases son: 10.2, 9.7, 10.1, 10.3, 10.1, 9.8, 9.9, 10.4, 10.3 y 9.8 litros. Utilice un nivel de significancia de 0.01 y suponga que la distribución del contenido es normal.

10.24 La estatura promedio de mujeres en el grupo de primer año de cierta universidad ha sido, históricamente, de 162.5 centímetros, con una desviación estándar de 6.9 centímetros. ¿Existe alguna razón para creer que ha habido un cambio en la estatura promedio, si una muestra aleatoria de 50 mujeres del grupo actual de primer año tiene una estatura promedio de 165.2 centímetros? Utilice un valor *P* en su conclusión. Suponga que la desviación estándar permanece constante.

10.25 Se afirma que los automóviles recorren en promedio más de 20,000 kilómetros por año. Para probar tal afirmación se pide a una muestra de 100 propietarios de automóviles seleccionada de manera aleatoria que lleven un registro de los kilómetros que recorren. ¿Estaría usted de acuerdo con esta afirmación, si la muestra aleatoria indicara un promedio de 23,500 kilómetros y una desviación estándar de 3900 kilómetros? Utilice un valor *P* en su conclusión.

10.26 De acuerdo con un estudio sobre un régimen alimenticio, la ingesta elevada de sodio se relaciona con úlceras, cáncer estomacal y migrañas. El requerimiento humano de sal es de tan sólo 220 miligramos diarios, el cual se rebasa en la mayoría de las porciones individuales de cereales listos para comerse. Si una muestra aleatoria de 20 porciones similares de cierto cereal tiene un contenido medio de 244 miligramos de sodio y una desviación estándar de 24.5 miligramos, ¿esto sugiere, a un nivel de significancia de 0.05, que el contenido promedio de sodio para porciones individuales de ese cereal es mayor que 220 miligramos? Suponga que la distribución de contenidos de sodio es normal.

10.27 Un estudio de la Universidad de Colorado en Boulder revela que correr aumenta el porcentaje de la tasa metabólica basal (TMB) en mujeres ancianas. La TMB promedio de 30 ancianas corredoras fue 34.0%

más alta que la TMB promedio de 30 ancianas sedentarias, en tanto que las desviaciones estándar reportadas fueron de 10.5 y 10.2%, respectivamente. ¿Existe un aumento significativo en la TMB de las corredoras respecto a las sedentarias? Suponga que las poblaciones se distribuyen de forma aproximadamente normal con varianzas iguales. Utilice un valor *P* en sus conclusiones.

10.28 De acuerdo con *Chemical Engineering,* una propiedad importante de la fibra es su absorbencia de agua. Se encontró que el porcentaje promedio de absorción de 25 pedazos de fibra de algodón seleccionados al azar es 20, con una desviación estándar de 1.5. Una muestra aleatoria de 25 pedazos de acetato reveló un porcentaje promedio de 12 con una desviación estándar de 1.25. ¿Existe evidencia sólida de que el porcentaje promedio de absorción de la población es significativamente mayor para la fibra de algodón que para el acetato? Suponga que el porcentaje de absorbencia se distribuye de forma casi normal y que las varianzas de la población en el porcentaje de absorbencia para las dos fibras son iguales. Utilice un nivel de significancia de 0.05.

10.29 La experiencia indica que el tiempo que requieren los estudiantes de último año de preparatoria para contestar una prueba estandarizada es una variable aleatoria normal con una media de 35 minutos. Si a una muestra aleatoria de 20 estudiantes de último año de preparatoria le toma un promedio de 33.1 minutos contestar esa prueba con una desviación estándar de 4.3 minutos, pruebe la hipótesis de que, a un nivel de significancia de 0.05, $\mu = 35$ minutos, contra la alternativa de que $\mu < 35$ minutos.

10.30 Una muestra aleatoria de tamaño $n_1 = 25$, tomada de una población normal con una desviación estándar $\sigma_1 = 5.2$, tiene una media $\bar{x}_1 = 81$. Una segunda muestra aleatoria de tamaño $n_2 = 36$, que se toma de una población normal diferente con una desviación estándar $\sigma_2 = 3.4$, tiene una media $\bar{x}_2 = 76$. Pruebe la hipótesis de que $\mu_1 = \mu_2$ contra la alternativa $\mu_1 \neq \mu_2$. Cite un valor *P* en su conclusión.

10.31 Un fabricante afirma que la resistencia promedio a la tensión del hilo *A* excede a la resistencia a la tensión promedio del hilo *B* en al menos 12 kilogramos. Para probar esta afirmación se pusieron a prueba 50 pedazos de cada tipo de hilo en condiciones similares. El hilo tipo *A* tuvo una resistencia promedio a la tensión de 86.7 kilogramos con una desviación estándar de 6.28 kilogramos; mientras que el hilo tipo *B* tuvo una resistencia promedio a la tensión de 77.8 kilogramos con una desviación estándar de 5.61 kilogramos. Pruebe la afirmación del fabricante usando un nivel de significancia de 0.05.

10.32 El *Amstat News* (diciembre de 2004) lista los sueldos medios de profesores asociados de estadística en instituciones de investigación, en escuelas de humanidades y en otras instituciones en Estados Unidos. Suponga que una muestra de 200 profesores asociados de instituciones de investigación tiene un sueldo promedio de \$70,750 anuales con una desviación estándar de \$6000. Suponga también que una muestra de 200 profesores asociados de otros tipos de instituciones tienen un sueldo promedio de \$65,200 con una desviación estándar de \$5000. Pruebe la hipótesis de que el sueldo medio de profesores asociados de instituciones de investigación es \$2000 más alto que el de los profesores de otras instituciones. Utilice un nivel de significancia de 0.01.

10.33 Se llevó a cabo un estudio para saber si el aumento en la concentración de sustrato tiene un efecto apreciable sobre la velocidad de una reacción química. Con una concentración de sustrato de 1.5 moles por litro, la reacción se realizó 15 veces, con una velocidad promedio de 7.5 micromoles por 30 minutos y una desviación estándar de 1.5. Con una concentración de sustrato de 2.0 moles por litro, se realizaron 12 reacciones que produjeron una velocidad promedio de 8.8 micromoles por 30 minutos y una desviación estándar muestral de 1.2. ¿Hay alguna razón para creer que este incremento en la concentración de sustrato ocasiona un aumento en la velocidad media de la reacción de más de 0.5 micromoles por 30 minutos? Utilice un nivel de significancia de 0.01 y suponga que las poblaciones se distribuyen de forma aproximadamente normal con varianzas iguales.

10.34 Se realizó un estudio para determinar si los temas de un curso de física se comprenden mejor cuando éste incluye prácticas de laboratorio. Se seleccionaron estudiantes al azar para que participaran en un curso de tres semestres con una hora de clase sin prácticas de laboratorio o en un curso de cuatro semestres con una hora de clase con prácticas de laboratorio. En la sección con prácticas de laboratorio 11 estudiantes obtuvieron una calificación promedio de 85 con una desviación estándar de 4.7; mientras que en la sección sin prácticas de laboratorio 17 estudiantes obtuvieron una calificación promedio de 79 con una desviación estándar de 6.1. ¿Diría usted que el curso que incluyó prácticas de laboratorio aumentó la calificación promedio hasta en 8 puntos? Utilice un valor *P* en su conclusión y suponga que las poblaciones se distribuyen de forma aproximadamente normal con varianzas iguales.

10.35 Para indagar si un nuevo suero frena el desarrollo de la leucemia se seleccionan 9 ratones, todos en una etapa avanzada de la enfermedad. Cinco ratones reciben el tratamiento y cuatro no. Los tiempos de supervivencia, en años, a partir del momento en que comienza el experimento son los siguientes:

Con tratamiento	2.1	5.3	1.4	4.6	0.9
Sin tratamiento	1.9	0.5	2.8	3.1	

A un nivel de significancia de 0.05, ¿se puede decir que el suero es eficaz? Suponga que las dos poblaciones se distribuyen de forma normal con varianzas iguales.

10.36 Los ingenieros de una armadora de automóviles de gran tamaño están tratando de decidir si comprarán neumáticos de la marca *A* o de la marca *B* para sus modelos nuevos. Con el fin de ayudarlos a tomar una decisión se realiza un experimento en el que se usan 12 neumáticos de cada marca. Los neumáticos se utilizan hasta que se desgastan. Los resultados son los siguientes:

Marca *A*:

$$\bar{x}_1 = 37{,}900 \text{ kilómetros},$$
$$s_1 = 5100 \text{ kilómetros}.$$

Marca *B*:

$$\bar{x}_1 = 39{,}800 \text{ kilómetros},$$
$$s_2 = 5900 \text{ kilómetros}.$$

Pruebe la hipótesis de que no hay diferencia en el desgaste promedio de las 2 marcas de neumáticos. Suponga que las poblaciones se distribuyen de forma aproximadamente normal con varianzas iguales. Use un valor *P*.

10.37 En el ejercicio 9.42 de la página 295 pruebe la hipótesis de que el ahorro de combustible de los camiones compactos Volkswagen, en promedio, excede al de los camiones compactos Toyota equipados de forma similar, que utilizan 4 kilómetros por litro. Utilice un nivel de significancia de 0.10.

10.38 Un investigador de la UCLA afirma que el promedio de vida de los ratones se puede prolongar hasta por 8 meses cuando se reducen las calorías en su dieta aproximadamente 40% desde el momento en que se destetan. Las dietas restringidas se enriquecen a niveles normales con vitaminas y proteínas. Suponga que a una muestra aleatoria de 10 ratones que tienen una vida promedio de 32.1 meses con una desviación estándar de 3.2 meses se les alimenta con una dieta normal, mientras que a una muestra aleatoria de 15 ratones que tienen un promedio de vida de 37.6 meses con una desviación estándar de 2.8 meses se les alimenta con la dieta restringida. A un nivel de significancia de 0.05 pruebe la hipótesis de que el promedio de vida de los ratones con esta dieta restringida aumenta 8 meses, contra la alternativa de que el aumento es menor de 8 meses. Suponga que las distribuciones de la esperanza de vida con las dietas regular y restringida son aproximadamente normales con varianzas iguales.

10.39 Los siguientes datos representan los tiempos de duración de películas producidas por 2 empresas cinematográficas:

Empresa	Tiempo (minutos)						
1	102	86	98	109	92		
2	81	165	97	134	92	87	114

Pruebe la hipótesis de que la duración promedio de las películas producidas por la empresa 2 excede al tiempo promedio de duración de las que produce la empresa 1 en 10 minutos, contra la alternativa unilateral de que la diferencia es de menos de 10 minutos. Utilice un nivel de significancia de 0.1 y suponga que las distribuciones de la duración son aproximadamente normales con varianzas iguales.

10.40 En un estudio realizado en Virginia Tech se compararon los niveles de ácido ascórbico en plasma en mujeres embarazadas fumadoras con los de mujeres no fumadoras. Para el estudio se seleccionaron 32 mujeres que estuvieran en los últimos 3 meses de embarazo, que no tuvieran padecimientos importantes y que sus edades fluctuaran entre los 15 y los 32 años. Antes de tomar muestras de 20 ml de sangre se pidió a las participantes que fueran en ayunas, que no tomaran sus suplementos vitamínicos y que evitaran alimentos con alto contenido de ácido ascórbico. A partir de las muestras de sangre se determinaron los siguientes valores de ácido ascórbico en el plasma de cada mujer, en miligramos por 100 mililitros:

Valores de ácido ascórbico en plasma

No fumadoras		Fumadoras
0.97	1.16	0.48
0.72	0.86	0.71
1.00	0.85	0.98
0.81	0.58	0.68
0.62	0.57	1.18
1.32	0.64	1.36
1.24	0.98	0.78
0.99	1.09	1.64
0.90	0.92	
0.74	0.78	
0.88	1.24	
0.94	1.18	

¿Existe suficiente evidencia para concluir que hay una diferencia entre los niveles de ácido ascórbico en plasma de mujeres fumadoras y no fumadoras? Suponga que los dos conjuntos de datos provienen de poblaciones normales con varianzas diferentes. Utilice un valor *P*.

10.41 El Departamento de Zoología de Virginia Tech llevó a cabo un estudio para determinar si existe una diferencia significativa en la densidad de organismos en dos estaciones diferentes ubicadas en Cedar Run, una corriente secundaria que se localiza en la cuenca del río Roanoke. El drenaje de una planta de tratamiento de aguas negras y el sobreflujo del estanque de sedimentación de la Federal Mogul Corporation entran al flujo cerca del nacimiento del río. Los siguientes datos proporcionan las medidas de densidad, en número de organismos por metro cuadrado, en las dos estaciones colectoras:

Número de organismos por metro cuadrado

Estación 1		Estación 2	
5030	4980	2800	2810
13,700	11,910	4670	1330
10,730	8130	6890	3320
11,400	26,850	7720	1230
860	17,660	7030	2130
2200	22,800	7330	2190
4250	1130		
15,040	1690		

A un nivel de significancia de 0.05, ¿podemos concluir que las densidades promedio en las dos estaciones son iguales? Suponga que las observaciones provienen de poblaciones normales con varianzas diferentes.

10.42 Cinco muestras de una sustancia ferrosa se usaron para determinar si existe una diferencia entre un análisis químico de laboratorio y un análisis de fluorescencia de rayos X del contenido de hierro. Cada muestra se dividió en dos submuestras y se aplicaron los dos tipos de análisis. A continuación se presentan los datos codificados que muestran los análisis de contenido de hierro:

	Muestra				
Análisis	**1**	**2**	**3**	**4**	**5**
Rayos X	2.0	2.0	2.3	2.1	2.4
Químico	2.2	1.9	2.5	2.3	2.4

Suponga que las poblaciones son normales y pruebe, al nivel de significancia de 0.05, si los dos métodos de análisis dan, en promedio, el mismo resultado.

10.43 De acuerdo con informes publicados, el ejercicio en condiciones de fatiga altera los mecanismos que determinan el desempeño. Se realizó un experimento con 15 estudiantes universitarios hombres, entrenados para realizar un movimiento horizontal continuo del brazo, de derecha a izquierda, desde un microinterruptor hasta una barrera, golpeando sobre la barrera en coincidencia con la llegada de una manecilla del reloj a la posición de las 6 en punto. Se registró el valor absoluto de la diferencia entre el tiempo, en milisegundos, que toma golpear sobre la barrera y el tiempo para que la manecilla alcance la posición de las 6 en punto (500 mseg). Cada participante ejecutó la tarea cinco veces en condiciones sin fatiga y con fatiga, y se registraron las siguientes sumas de las diferencias absolutas para las cinco ejecuciones:

	Diferencias absolutas de tiempo	
Sujeto	**Sin fatiga**	**Con fatiga**
1	158	91
2	92	59
3	65	215
4	98	226
5	33	223
6	89	91
7	148	92
8	58	177
9	142	134
10	117	116
11	74	153
12	66	219
13	109	143
14	57	164
15	85	100

Un aumento en la diferencia media absoluta de tiempo cuando la tarea se ejecuta en condiciones de fatiga apoyaría la afirmación de que el ejercicio, en condiciones de fatiga, altera el mecanismo que determina el desempeño. Suponga que las poblaciones se distribuyen normalmente y pruebe tal afirmación.

10.44 En un estudio realizado por el Departamento de Nutrición Humana y Alimentos del Virginia Tech se registraron los siguientes datos sobre los residuos de ácido sórbico en jamón, en partes por millón, inmediatamente después de sumergirlo en una solución de sorbato y después de 60 días de almacenamiento:

	Residuos de ácido sórbico en jamón	
Rebanada	**Antes del almacenamiento**	**Después del almacenamiento**
1	224	116
2	270	96
3	400	239
4	444	329
5	590	437
6	660	597
7	1400	689
8	680	576

Si se supone que las poblaciones se distribuyen normalmente, ¿hay suficiente evidencia, a un nivel de significancia de 0.05, para decir que la duración del almacenamiento influye en las concentraciones residuales de ácido sórbico?

10.45 El administrador de una empresa de taxis está tratando de decidir si el uso de neumáticos radiales en lugar de neumáticos regulares cinturados mejora el rendimiento de combustible. Se equipan 12 autos con neumáticos radiales y se conducen en un recorrido de prueba preestablecido. Sin cambiar a los conductores, los mismos autos se equipan con neumáticos regulares cinturados y se conducen nuevamente en el recorrido de prueba. Se registraron los siguientes datos sobre el consumo de gasolina, en kilómetros por litro:

	Kilómetros por litro	
Automóvil	**Llantas radiales**	**Llantas cinturadas**
1	4.2	4.1
2	4.7	4.9
3	6.6	6.2
4	7.0	6.9
5	6.7	6.8
6	4.5	4.4
7	5.7	5.7
8	6.0	5.8
9	7.4	6.9
10	4.9	4.7
11	6.1	6.0
12	5.2	4.9

¿Podemos concluir que los autos equipados con neumáticos radiales ahorran más combustible que aquellos equipados con neumáticos cinturados? Suponga que las poblaciones se distribuyen normalmente. Utilice un valor P en su conclusión.

10.46 En el ejercicio de repaso 9.91 de la página 313 utilice la distribución t para probar la hipótesis de que la dieta reduce el peso de un individuo en 4.5 kilogramos, en promedio, contra la hipótesis alternativa de que la diferencia media en peso es menor que 4.5 kilogramos. Utilice un valor P.

10.47 ¿Qué tan grande debería ser la muestra del ejercicio 10.20 para que la potencia de la prueba sea de 0.90, cuando la media verdadera es 5.20? Suponga que $\sigma = 0.24$.

10.48 Si la distribución del tiempo de vida en el ejercicio 10.19 es aproximadamente normal, ¿qué tan grande debería ser una muestra para que la probabilidad de cometer un error tipo II sea 0.1 cuando la media verdadera es 35.9 meses? Suponga que $\sigma = 5.8$ meses.

10.49 ¿Qué tan grande debería ser la muestra del ejercicio 10.24 para que la potencia de la prueba sea de 0.95 cuando la estatura promedio verdadera difiere de 162.5 en 3.1 centímetros? Utilice $\alpha = 0.02$.

10.50 ¿Qué tan grandes deberían ser las muestras del ejercicio 10.31 para que la potencia de la prueba sea de 0.95, cuando la diferencia verdadera entre los tipos de hilo A y B es 8 kilogramos?

10.51 ¿Qué tan grande debería ser la muestra del ejercicio 10.22 para que la potencia de la prueba sea de 0.8 cuando el tiempo promedio verdadero dedicado a la meditación excede al valor hipotético en 1.2 σ? Utilice $\alpha = 0.05$.

10.52 Se considera una prueba t a un nivel $\alpha = 0.05$ para probar

$$H_0\colon \mu = 14,$$
$$H_1\colon \mu \neq 14.$$

¿Qué tamaño de muestra se necesita para que la probabilidad de no rechazar de manera errónea H_0 sea 0.1 cuando la media de la población verdadera difiere de 14 en 0.5? A partir de una muestra preliminar estimamos que σ es 1.25.

10.53 En el Departamento de Medicina Veterinaria del Virginia Tech se llevó a cabo un estudio para determinar si la "resistencia" de una herida de incisión quirúrgica es afectada por la temperatura del bisturí. En el experimento se utilizaron 8 perros. Se hicieron incisiones "calientes" y "frías" en el abdomen de cada perro y se midió la resistencia. A continuación se presentan los datos resultantes.

Perro	Bisturí	Resistencia
1	Caliente	5120
1	Frío	8200
2	Caliente	10,000
2	Frío	8600
3	Caliente	10,000
3	Frío	9200
4	Caliente	10,000
4	Frío	6200
5	Caliente	10,000
5	Frío	10,000
6	Caliente	7900
6	Frío	5200
7	Caliente	510
7	Frío	885
8	Caliente	1020
8	Frío	460

a) Escriba una hipótesis adecuada para determinar si la resistencia de las incisiones realizadas con bisturí caliente difiere en forma significativa de la resistencia de las realizadas con bisturí frío.

b) Pruebe la hipótesis utilizando una prueba t pareada. Utilice un valor P en su conclusión.

10.54 Se utilizaron 9 sujetos en un experimento para determinar si la exposición a monóxido de carbono tiene un impacto sobre la capacidad respiratoria. Los datos fueron recolectados por el personal del Departamento de Salud y Educación Física del Virginia Tech y analizados en el Centro de Consulta Estadística en Hokie Land. Los sujetos fueron expuestos a cámaras de respiración, una de las cuales contenía una alta concentración de CO. Se realizaron varias mediciones de frecuencia respiratoria a cada sujeto en cada cámara. Los sujetos fueron expuestos a las cámaras de respiración en una secuencia aleatoria. Los siguientes datos representan la frecuencia respiratoria en número de respiraciones por minuto. Realice una prueba unilateral de la hipótesis de que la frecuencia respiratoria media es igual en los dos ambientes. Utilice $\alpha = 0.05$. Suponga que la frecuencia respiratoria es aproximadamente normal.

Sujeto	Con CO	Sin CO
1	30	30
2	45	40
3	26	25
4	25	23
5	34	30
6	51	49
7	46	41
8	32	35
9	30	28

10.8 Una muestra: prueba sobre una sola proporción

Las pruebas de hipótesis que se relacionan con proporciones se requieren en muchas áreas. A los políticos les interesa conocer la fracción de votantes que los favorecerá en la siguiente elección. Todas las empresas manufactureras se preocupan por la proporción de artículos defectuosos cuando se realiza un embarque. Los jugadores dependen del conocimiento de la proporción de resultados que consideran favorables.

Consideraremos el problema de probar la hipótesis de que la proporción de éxitos en un experimento binomial es igual a algún valor específico. Es decir, probaremos la hipótesis nula H_0 de que $p = p_0$, donde p es el parámetro de la distribución binomial. La hipótesis alternativa puede ser una de las alternativas unilaterales o bilaterales usuales:

$$p < p_0, \quad p > p_0, \quad \text{o} \quad p \neq p_0.$$

La variable aleatoria adecuada sobre la que basamos nuestro criterio de decisión es la variable aleatoria binomial X; aunque también podríamos usar el estadístico $\hat{p} = X/n$. Los valores de X que están lejos de la media $\mu = np_0$ conducirán al rechazo de la hipótesis nula. Como X es una variable binomial discreta, es poco probable que se pueda establecer una región crítica cuyo tamaño sea *exactamente* igual a un valor preestablecido de α. Por esta razón es preferible, al trabajar con muestras pequeñas, basar nuestras decisiones en valores P. Para probar la hipótesis

$$H_0\text{: } p = p_0,$$
$$H_1\text{: } p < p_0,$$

utilizamos la distribución binomial para calcular el valor P

$$P = P(X \leq x \text{ cuando } p = p_0).$$

El valor x es el número de éxitos en nuestra muestra de tamaño n. Si este valor P es menor o igual que α, nuestra prueba es significativa al nivel α y rechazamos H_0 a favor de H_1. De manera similar, para probar la hipótesis

$$H_0\text{: } p = p_0,$$
$$H_1\text{: } p > p_0,$$

al nivel de significancia α, calculamos

$$P = P(X \geq x \text{ cuando } p = p_0)$$

y rechazamos H_0 a favor de H_1 si este valor P es menor o igual que α. Finalmente, para probar la hipótesis

$$H_0\text{: } p = p_0,$$
$$H_1\text{: } p \neq p_0,$$

a un nivel de significancia α, calculamos

$$P = 2P(X \leq x \text{ cuando } p = p_0) \qquad \text{si } x < np_0$$

o

$$P = 2P(X \geq x \text{ cuando } p = p_0) \qquad \text{si } x > np_0$$

y rechazamos H_0 a favor de H_1 si el valor P calculado es menor o igual que α.

Los pasos para probar una hipótesis nula acerca de una proporción contra varias alternativas usando las probabilidades binomiales de la tabla A.1 son los siguientes:

Prueba de una proporción (muestras pequeñas)

1. H_0: $p = p_0$.
2. Una de las alternativas H_1: $p < p_0$, $p > p_0$ o $p \neq p_0$.
3. Elegir un nivel de significancia igual a α.
4. Estadístico de prueba: variable binomial X con $p = p_0$.
5. Cálculos: obtener x, el número de éxitos, y calcular el valor P adecuado.
6. Decisión: sacar las conclusiones apropiadas con base en el valor P.

Ejemplo 10.9: Un constructor afirma que en 70% de las viviendas que se construyen actualmente en la ciudad de Richmond, Virginia, se instalan bombas de calor. ¿Estaría de acuerdo con esta afirmación si una encuesta aleatoria de viviendas nuevas en esta ciudad revelara que 8 de 15 tienen instaladas bombas de calor? Utilice un nivel de significancia de 0.10.

***Solución*:**

1. H_0: $p = 0.7$.
2. H_1: $p \neq 0.7$.
3. $\alpha = 0.10$.
4. Estadístico de prueba: Variable binomial X con $p = 0.7$ y $n = 15$.
5. Cálculos: $x = 8$ y $np_0 = (15)(0.7) = 10.5$. Por lo tanto, de la tabla A.1, el valor P calculado es

$$P = 2P(X \leq 8 \text{ cuando } p = 0.7) = 2\sum_{x=0}^{8} b(x; 15, 0.7) = 0.2622 > 0.10.$$

6. Decisión: No rechazar H_0. Concluir que no hay razón suficiente para dudar de la afirmación del constructor.

En la sección 5.2 aprendimos que cuando n es pequeña las probabilidades binomiales se pueden obtener de la fórmula binomial real o de la tabla A.1. Para n grande se requieren procedimientos de aproximación. Cuando el valor hipotético p_0 está muy cerca de 0 o de 1 se puede utilizar la distribución de Poisson con parámetro $\mu = np_0$. Sin embargo, para n grande por lo general se prefiere la aproximación de la curva normal, con los parámetros $\mu = np_0$ y $\sigma^2 = np_0q_0$, la cual es muy precisa, siempre y cuando p_0 no esté demasiado cerca de 0 o de 1. Si utilizamos la aproximación normal, el **valor z para probar $p = p_0$** es dado por

$$z = \frac{x - np_0}{\sqrt{np_0q_0}} = \frac{\hat{p} - p_0}{\sqrt{p_0q_0/n}},$$

que es un valor de la variable normal estándar Z. Por consiguiente, para una prueba de dos colas al nivel de significancia α, la región crítica es $z < -z_{\alpha/2}$ o $z > z_{\alpha/2}$. Para la alternativa unilateral $p < p_0$, la región crítica es $z < -z_\alpha$, y para la alternativa $p > p_0$, la región crítica es $z > z_\alpha$.

Ejemplo 10.10: Se considera que un medicamento que se prescribe comúnmente para aliviar la tensión nerviosa tiene una eficacia de tan sólo 60%. Los resultados experimentales de un nuevo fármaco administrado a una muestra aleatoria de 100 adultos que padecían tensión nerviosa revelaron que 70 de ellos sintieron alivio. ¿Esta evidencia es suficiente para concluir que el nuevo medicamento es mejor que el que se prescribe comúnmente? Utilice un nivel de significancia de 0.05.

***Solución*:**

1. H_0: $p = 0.6$.
2. H_1: $p > 0.6$.
3. $\alpha = 0.05$.
4. Región crítica: $z > 1.645$.
5. Cálculos: $x = 70$, $n = 100$, $\hat{p} = 70/100 = 0.7$ y

$$z = \frac{0.7-0.6}{\sqrt{(0.6)(0.4)/100}} = 2.04,\ P = P(Z > 2.04) < 0.0207.$$

6. Decisión: Rechazar H_0 y concluir que el nuevo fármaco es mejor.

10.9 Dos muestras: pruebas sobre dos proporciones

A menudo surgen situaciones en las que se desea probar la hipótesis de que dos proporciones son iguales. Por ejemplo, podemos tratar de mostrar evidencia de que la proporción de médicos que son pediatras en un estado es igual a la proporción de pediatras en otro estado. Quizás un individuo decida dejar de fumar sólo si se convence de que la proporción de fumadores con cáncer pulmonar excede a la proporción de no fumadores con ese tipo de cáncer.

En general, deseamos probar la hipótesis nula de que dos proporciones, o parámetros binomiales, son iguales. Es decir, probamos $p_1 = p_2$ contra una de las alternativas $p_1 < p_2$, $p_1 > p_2$, o $p_1 \neq p_2$. Desde luego, esto es equivalente a probar la hipótesis nula de que $p_1 - p_2 = 0$ contra una de las alternativas $p_1 - p_2 < 0$, $p_1 - p_2 > 0$ o $p_1 - p_2 \neq 0$. El estadístico sobre el que basamos nuestra decisión es la variable aleatoria $\hat{P}_1 - \hat{P}_2$. Se seleccionan al azar muestras independientes de tamaños n_1 y n_2 de dos poblaciones binomiales y se calcula la proporción de éxitos $\hat{P}_1$ y $\hat{P}_2$ para las dos muestras.

En la construcción de intervalos de confianza para p_1 y p_2 observamos, para n_1 y n_2 suficientemente grandes, que el estimador puntual $\hat{P}_1$ menos $\hat{P}_2$ estaba distribuido de forma casi normal con media

$$\mu_{\hat{P}_1 - \hat{P}_2} = p_1 - p_2$$

y varianza

$$\sigma^2_{\hat{P}_1 - \hat{P}_2} = \frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}.$$

Por lo tanto, es posible establecer la(s) región(es) crítica(s) usando la variable normal estándar

$$Z = \frac{(\hat{P}_1 - \hat{P}_2) - (p_1 - p_2)}{\sqrt{p_1 q_1/n_1 + p_2 q_2/n_2}}.$$

Cuando H_0 es verdadera, podemos sustituir $p_1 = p_2 = p$ y $q_1 = q_2 = q$ (donde p y q son los valores comunes) en la fórmula anterior para Z y obtener la forma

$$Z = \frac{\hat{P}_1 - \hat{P}_2}{\sqrt{pq(1/n_1 + 1/n_2)}}.$$

Sin embargo, para calcular un valor de Z debemos estimar los parámetros p y q que aparecen en el radical. Al agrupar los datos de ambas muestras el **estimado agrupado de la proporción** p es

$$\hat{p} = \frac{x_1 + x_2}{n_1 + n_2},$$

donde x_1 y x_2 son el número de éxitos en cada una de las dos muestras. Al sustituir $\hat{p}$ por p y $\hat{q} = 1 - \hat{p}$ por q, el **valor z para probar $p_1 = p_2$** se determina a partir de la fórmula

$$z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}\hat{q}(1/n_1 + 1/n_2)}}.$$

Las regiones críticas para las hipótesis alternativas adecuadas se establecen como antes, utilizando puntos críticos de la curva normal estándar. En consecuencia, para la alternativa $p_1 \neq p_2$, al nivel de significancia α, la región crítica es $z < -z_{\alpha/2}$ o $z > z_{\alpha/2}$. Para una prueba donde la alternativa es $p_1 < p_2$, la región crítica será $z < -z_\alpha$; y cuando la alternativa es $p_1 > p_2$, la región crítica será $z > z_\alpha$.

Ejemplo 10.11: Se organizará una votación entre los residentes de una ciudad y el condado circundante para determinar si se aprueba una propuesta para la construcción de una planta química. Como el lugar en el que se propone construirla está dentro de los límites de la ciudad, muchos votantes del condado consideran que la propuesta será aprobada debido a la gran proporción de votantes que está a favor de que se construya. Se realiza una encuesta para determinar si hay una diferencia significativa en la proporción de votantes de la ciudad y los votantes del condado que favorecen la propuesta. Si 120 de 200 votantes de la ciudad favorecen la propuesta y 240 de 500 residentes del condado también lo hacen, ¿estaría usted de acuerdo en que la proporción de votantes de la ciudad que favorecen la propuesta es mayor que la proporción de votantes del condado? Utilice un nivel de significancia de $\alpha = 0.05$.

***Solución*:** Sean p_1 y p_2 las proporciones verdaderas de votantes en la ciudad y el condado, respectivamente, que favorecen la propuesta.

1. H_0: $p_1 = p_2$.

2. H_1: $p_1 > p_2$.

3. $\alpha = 0.05$

4. Región crítica: $z > 1.645$.

5. Cálculos:

$$\hat{p}_1 = \frac{x_1}{n_1} = \frac{120}{200} = 0.60, \quad \hat{p}_2 = \frac{x_2}{n_2} = \frac{240}{500} = 0.48, \quad \text{y}$$
$$\hat{p} = \frac{x_1 + x_2}{n_1 + n_2} = \frac{120 + 240}{200 + 500} = 0.51.$$

Por lo tanto,

$$z = \frac{0.60 - 0.48}{\sqrt{(0.51)(0.49)(1/200 + 1/500)}} = 2.9,$$
$$P = P(Z > 2.9) = 0.0019.$$

6. Decisión: Rechazar H_0 y estar de acuerdo en que la proporción de votantes de la ciudad a favor de la propuesta es mayor que la proporción de votantes del condado.

Ejercicios

10.55 Un experto en mercadotecnia de una empresa fabricante de pasta considera que 40% de los amantes de la pasta prefieren la lasagna. Si 9 de 20 amantes de la pasta eligen la lasagna sobre otras pastas, ¿qué se puede concluir acerca de la afirmación del experto? Utilice un nivel de significancia de 0.05.

10.56 Suponga que, en el pasado, 40% de todos los adultos estaban a favor de la pena capital. ¿Existe alguna razón para creer que la proporción de adultos que está a favor de la pena capital ha aumentado si, en una muestra aleatoria de 15 adultos, 8 están a favor de la pena capital? Utilice un nivel de significancia de 0.05.

10.57 Se está considerando utilizar un nuevo aparato de radar para cierto sistema de misiles de defensa. El sistema se verifica experimentando con una aeronave en la que se simula una situación en la que alguien muere y otra en la que no ocurre ninguna muerte. Si en 300 ensayos ocurren 250 muertes, al nivel de significancia de 0.04, acepte o rechace la afirmación de que la probabilidad de una muerte con el nuevo sistema no excede a la probabilidad de 0.8 del sistema que se utiliza actualmente.

10.58 Se cree que al menos 60% de los residentes de cierta área están a favor de una demanda de anexión de una ciudad vecina. ¿Qué conclusión extraería si sólo 110 en una muestra de 200 votantes están a favor de la demanda? Utilice un nivel de significancia de 0.05.

10.59 Una empresa petrolera afirma que en una quinta parte de las viviendas de cierta ciudad la gente utiliza petróleo como combustible para calentarlas. ¿Existen razones para creer que en menos de una quinta parte de las viviendas la gente utiliza este combustible para calentarlas si, en una muestra aleatoria de 1000 viviendas de esa ciudad, se encuentra que 136 utilizan petróleo como combustible? Utilice un valor P en su conclusión.

10.60 En cierta universidad se estima que a lo sumo 25% de los estudiantes van en bicicleta a la escuela. ¿Parece que ésta es una estimación válida si, en una muestra aleatoria de 90 estudiantes universitarios, se encuentra que 28 van en bicicleta a la escuela? Utilice un nivel de significancia de 0.05.

10.61 En un invierno con epidemia de influenza los investigadores de una conocida empresa farmacéutica encuestaron a los padres de 2000 bebés para determinar si el nuevo medicamento de la empresa era eficaz después de dos días. De 120 bebés que tenían influenza y que recibieron el medicamento, 29 se curaron en dos días o menos. De 280 bebés que tenían influenza pero no recibieron el fármaco, 56 se curaron en dos días o menos. ¿Hay alguna indicación significativa que apoye la afirmación de la empresa sobre la eficacia del medicamento?

10.62 En un experimento de laboratorio controlado, científicos de la Universidad de Minnesota descubrieron que 25% de cierta cepa de ratas sujetas a una dieta con 20% de grano de café y luego forzadas a consumir un poderoso químico causante de cáncer desarrollaron tumores cancerosos. Si el experimento se repite, y 16 de 48 ratas desarrollan tumores, ¿existen razones para creer que la proporción de ratas que desarrollan tumores cuando se someten a esta dieta se incrementa? Utilice un nivel de significancia de 0.05.

10.63 En un estudio que se realizó para estimar la proporción de residentes de cierta ciudad y sus suburbios que están a favor de que se construya una planta de energía nuclear se encontró que 63 de 100 residentes urbanos están a favor de la construcción, mientras que sólo 59 de 125 residentes suburbanos la apoyan. ¿Hay una diferencia significativa entre la proporción de residentes urbanos y suburbanos que están a favor de que se construya la planta nuclear? Utilice un valor *P*.

10.64 En un estudio sobre la fertilidad de mujeres casadas, realizado por Martin O'Connell y Carolyn C. Rogers para la Oficina del Censo en 1979, se seleccionaron al azar dos grupos de mujeres casadas de entre 25 y 29 años de edad y sin hijos, y a cada una se le preguntó si planeaba tener un hijo en algún momento. Se seleccionó un grupo de mujeres con menos de dos años de casadas y otro de mujeres con cinco años de casadas. Suponga que 240 de 300 mujeres con menos de dos años de casadas planean tener un hijo algún día, en comparación con 288 de las 400 mujeres con cinco años de casadas. ¿Podemos concluir que la proporción de mujeres con menos de dos años de casadas que planean tener hijos es significativamente mayor que la proporción de mujeres con cinco años de casadas que también planean tenerlos? Utilice un valor *P*.

10.65 Una comunidad urbana quiere demostrar que la incidencia de cáncer de mama es mayor en su localidad que en una área rural vecina. (Se encontró que los niveles de PCB son más altos en el suelo de la comunidad urbana). Si descubre que en la comunidad urbana 20 de 200 mujeres adultas tienen cáncer de mama y que en la comunidad rural 10 de 150 mujeres adultas lo tienen, ¿podría concluir, con un nivel de significancia de 0.05, que el cáncer de mama prevalece más en la comunidad urbana?

10.66 **Proyecto de grupo:** Para este proyecto el grupo se debe dividir en parejas. Suponga que se supone que al menos 25% de los estudiantes de su universidad hacen más de dos horas de ejercicio por semana. Reúna datos de una muestra aleatoria de 50 estudiantes y pregunte a cada uno si se ejercita durante al menos dos horas por semana; luego haga los cálculos necesarios para rechazar o no rechazar la suposición anterior. Demuestre todo el procedimiento y utilice un valor *P* en sus conclusiones.

10.10 Pruebas de una y dos muestras referentes a varianzas

En esta sección estudiaremos la prueba de hipótesis relacionada con varianzas o desviaciones estándar de la población. No son poco comunes las aplicaciones de pruebas de una y dos muestras sobre varianzas. Los ingenieros y los científicos constantemente se enfrentan a estudios donde se les pide demostrar que las mediciones que tienen que ver con productos o procesos cumplen con las especificaciones que fijan los consumidores. Las especificaciones a menudo se cumplen si la varianza del proceso es suficientemente pequeña. También existe interés por experimentos que comparan métodos o procesos donde la reproducibilidad o variabilidad inherentes se deben comparar de manera formal. Además, para determinar si no se cumple la suposición de varianzas iguales, con frecuencia se aplica una prueba que compara dos varianzas antes de llevar a cabo una prueba *t* sobre dos medias.

Empecemos por considerar el problema de probar la hipótesis nula H_0 de que la varianza de la población σ^2 es igual a un valor específico σ_0^2 contra una de las alternativas comunes $\sigma^2 < \sigma_0^2$, $\sigma^2 > \sigma_0^2$ o $\sigma^2 \neq \sigma_0^2$. El estadístico apropiado sobre el que basamos nuestra decisión es el estadístico chi cuadrada del teorema 8.4, el cual se utilizó en el capítulo 9 para construir un intervalo de confianza para σ^2. Por lo tanto, si suponemos que la distribución de la población que se muestrea es normal, el valor de chi cuadrada para probar $\sigma^2 = \sigma_0^2$ es dado por

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

donde n es el tamaño de la muestra, s^2 es la varianza muestral y σ_0^2 es el valor de σ^2 dado por la hipótesis nula. Si H_0 es verdadera, χ^2 es un valor de la distribución chi cuadrada con $v = n - 1$ grados de libertad. En consecuencia, para una prueba de dos colas a un nivel de significancia α, la región crítica es $\chi^2 < \chi^2_{1-\alpha/2}$ o $\chi^2 > \chi^2_{\alpha/2}$. Para la alternativa unilateral $\sigma^2 < \sigma_0^2$, la región crítica es $\chi^2 < \chi^2_{1-\alpha}$; y para la alternativa unilateral $\sigma^2 > \sigma_0^2$, la región crítica es $\chi^2 > \chi^2_{\alpha}$.

Robustez de la prueba χ^2 para la suposición de normalidad

Tal vez el lector se habrá dado cuenta de que varias pruebas dependen, al menos en teoría, de la suposición de normalidad. En general muchos procedimientos en estadística aplicada tienen fundamentos teóricos que dependen de la distribución normal. Estos procedimientos varían en el grado en que dependen de la suposición de la normalidad. A un procedimiento que es razonablemente insensible a esta suposición se le denomina **procedimiento robusto**, es decir, robusto para la normalidad. La prueba χ^2 sobre una sola varianza no es robusta en absoluto para la normalidad, es decir, el éxito práctico del procedimiento depende de la normalidad. Como resultado, el valor *P* calculado podría ser notoriamente diferente del valor *P* verdadero si la población de la que se toma la muestra no es normal. De hecho, resulta muy plausible que un valor *P* estadísticamente significativo no sea una verdadera señal de H_1: $\sigma \neq \sigma_0$, sino que un valor significativo sea el resultado de haber violado las suposiciones de normalidad. Por lo tanto, el analista debería utilizar esta prueba χ^2 específica con precaución.

Ejemplo 10.12: Un fabricante de baterías para automóvil afirma que la duración de sus baterías se distribuye de forma aproximadamente normal con una desviación estándar igual a 0.9 años. Si una muestra aleatoria de 10 de tales baterías tiene una desviación estándar de 1.2 años, ¿considera que $\sigma > 0.9$ años? Utilice un nivel de significancia de 0.05.

***Solución*:**

1. H_0: $\sigma^2 = 0.81$.
2. H_1: $\sigma^2 > 0.81$.
3. $\alpha = 0.05$.
4. Región crítica: En la figura 10.19 vemos que se rechaza la hipótesis nula cuando $\chi^2 > 16.919$, donde $\chi^2 = \frac{(n-1)s^2}{\sigma_0^2}$ con $v = 9$ grados de libertad.

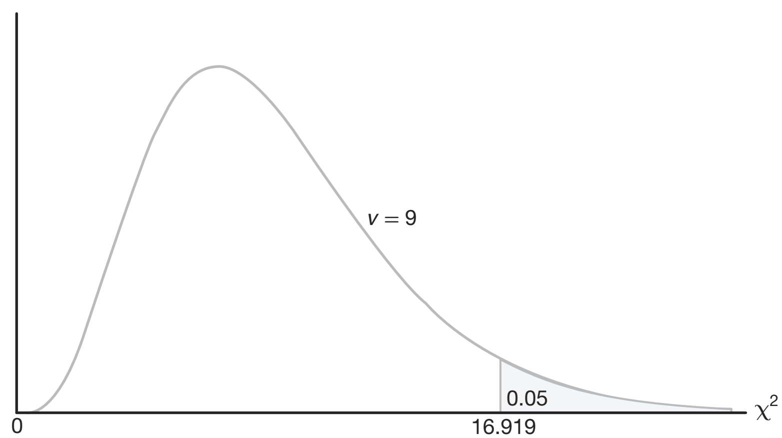

Figura 10.19: Región crítica para la hipótesis alternativa $\sigma > 0.9$.

5. Cálculos: $s^2 = 1.44$, $n = 10$ y

$$\chi^2 = \frac{(9)(1.44)}{0.81} = 16.0, \qquad P \approx 0.07.$$

6. Decisión: El estadístico χ^2 no es significativo al nivel 0.05. Sin embargo, con base en el valor P de 0.07, hay evidencia de que $\sigma > 0.9$.

Consideremos ahora el problema de probar la igualdad de las varianzas σ_1^2 y σ_2^2 de dos poblaciones. Esto es, probaremos la hipótesis nula H_0 de que $\sigma_1^2 = \sigma_2^2$ contra una de las alternativas usuales

$$\sigma_1^2 < \sigma_2^2, \quad \sigma_1^2 > \sigma_2^2, \quad \text{o} \quad \sigma_1^2 \neq \sigma_2^2.$$

Para muestras aleatorias independientes de tamaños n_1 y n_2, respectivamente, de las dos poblaciones, **el valor f para probar σ = σ** es el cociente

$$f = \frac{s_1^2}{s_2^2},$$

donde s_1^2 y s_2^2 son las varianzas calculadas de las dos muestras. Si las dos poblaciones se distribuyen de forma aproximadamente normal y la hipótesis nula es verdadera, de acuerdo con el teorema 8.8 el cociente $f = s_1^2 / s_2^2$ es un valor de la distribución F con $v_1 = n_1 - 1$ y $v_2 = n_2 - 1$ grados de libertad. Por lo tanto, las regiones críticas de tamaño α que corresponden a las alternativas unilaterales $\sigma_1^2 < \sigma_2^2$ y $\sigma_1^2 > \sigma_2^2$ son, respectivamente, $f < f_{1-\alpha}(v_1, v_2)$ y $f > f_\alpha(v_1, v_2)$. Para la alternativa bilateral $\sigma_1^2 \neq \sigma_2^2$ la región crítica es $f < f_{1-\alpha/2}(v_1, v_2)$ o $f > f_{\alpha/2}(v_1, v_2)$.

Ejemplo 10.13: Al probar la diferencia en el desgaste abrasivo de los dos materiales del ejemplo 10.6 supusimos que las dos varianzas de la población desconocidas eran iguales. ¿Se justifica tal suposición? Utilice un nivel de significancia de 0.10.

***Solución*:** Sean σ_1^2 y σ_2^2 las varianzas de la población para el desgaste abrasivo del material 1 y del material 2, respectivamente.

1. H_0: $\sigma_1^2 = \sigma_2^2$
2. H_1: $\sigma_1^2 \neq \sigma_2^2$
3. $\alpha = 0.10$.
4. Región crítica: En la figura 10.20 observamos que $f_{0.05}(11, 9) = 3.11$, y, usando el teorema 8.7, encontramos

$$f_{0.95}(11, 9) = \frac{1}{f_{0.05}(9, 11)} = 0.34.$$

Por lo tanto, se rechaza la hipótesis nula cuando $f < 0.34$ o $f > 3.11$, donde $f = s_1^2 / s_2^2$ con $\nu_1 = 11$ y $\nu_2 = 9$ grados de libertad.

5. Cálculos: $s_1^2 = 16$, $s_2^2 = 25$, por ende, $f = \frac{16}{25} = 0.64$.
6. Decisión: no rechazar H_0. Concluir que no hay suficiente evidencia de que las varianzas sean diferentes.

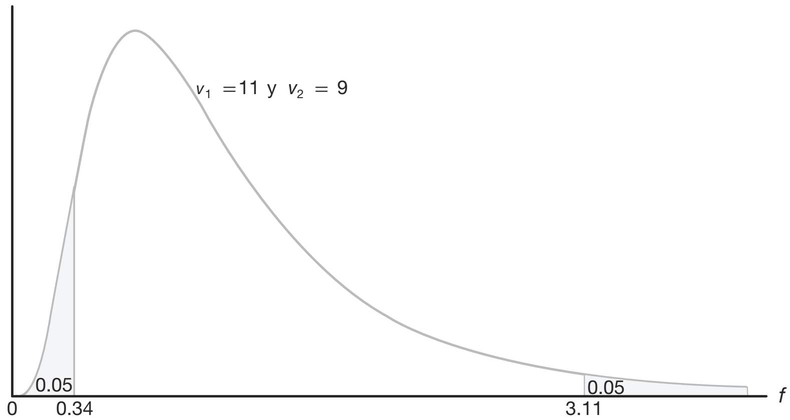

Figura 10.20: Región crítica para la hipótesis alternativa $\sigma_1^2 \neq \sigma_2^2$.

Prueba *F* para la prueba de varianzas con el SAS

La figura 10.18 de la página 356 presenta la impresión de una prueba *t* de dos muestras donde se comparan dos medias de los datos de los tallos en el ejercicio 9.40. La gráfica de caja y bigote que se observa en la figura 10.17 de la página 355 sugiere que las varianzas no son homogéneas y, por consiguiente, el estadístico t' y su valor *P* correspondiente son relevantes. Observe también que la impresión muestra el estadístico *F* para H_0: $\sigma_1 = \sigma_2$ con un valor *P* de 0.0098, que es evidencia adicional de que se debe esperar más variabilidad cuando se aplica el tratamiento con nitrógeno que cuando no se aplica.

Ejercicios

10.67 Se sabe que el contenido de los envases de un lubricante específico se distribuye normalmente con una varianza de 0.03 litros. Pruebe la hipótesis de que $\sigma^2 = 0.03$ contra la alternativa de que $\sigma^2 \neq 0.03$ para la muestra aleatoria de 10 envases del ejercicio 10.23 de la página 356. Use un valor *P* en sus conclusiones.

10.68 Por experiencia se sabe que el tiempo que se requiere para que los estudiantes de preparatoria de último año contesten una prueba estandarizada es una variable aleatoria normal con una desviación estándar de 6 minutos. Pruebe la hipótesis de que $\sigma = 6$ contra la alternativa de que $\sigma < 6$ si una muestra aleatoria de los tiempos para realizar la prueba de 20 estudiantes de preparatoria de último año tiene una desviación estándar $s = 4.51$. Utilice un nivel de significancia de 0.05.

10.69 Se deben supervisar las aflotoxinas ocasionadas por moho en cosechas de cacahuate en Virginia. Una muestra de 64 lotes de cacahuate revela niveles de 24.17 ppm, en promedio, con una varianza de 4.25 ppm. Pruebe la hipótesis de que $\sigma^2 = 4.2$ ppm contra la alternativa de que $\sigma^2 \neq 4.2$ ppm. Utilice un valor *P* en sus conclusiones.

10.70 Datos históricos indican que la cantidad de dinero que aportaron los residentes trabajadores de una ciudad grande para un escuadrón de rescate voluntario es una variable aleatoria normal con una desviación estándar de \$1.40. Se sugiere que las contribuciones al escuadrón de rescate sólo de los empleados del departamento de sanidad son mucho más variables. Si las contribuciones de una muestra aleatoria de 12 empleados del departamento de sanidad tienen una desviación estándar de \$1.75, ¿podemos concluir a un nivel de significancia de 0.01 que la desviación estándar de las contribuciones de todos los trabajadores de sanidad es mayor que la de todos los trabajadores que viven en dicha ciudad?

10.71 Se dice que una máquina despachadora de bebida gaseosa está fuera de control si la varianza de los contenidos excede a 1.15 decilitros. Si una muestra aleatoria de 25 bebidas de esta máquina tiene una varianza de 2.03 decilitros, ¿esto indica, a un nivel de significancia de 0.05, que la máquina está fuera de control? Suponga que los contenidos se distribuyen de forma aproximadamente normal.

10.72 Prueba de $\sigma^2 = \sigma_0^2$ para una muestra grande: Cuando $n \geq 30$ podemos probar la hipótesis nula de que $\sigma^2 = \sigma_0^2$ o $\sigma = \sigma_0$ calculando

$$z = \frac{s - \sigma_0}{\sigma_0/\sqrt{2n}},$$

que es un valor de una variable aleatoria cuya distribución muestral es aproximadamente la distribución normal estándar.

a) Con referencia al ejemplo 10.4, a un nivel de significancia de 0.05, pruebe si $\sigma = 10.0$ años contra la alternativa de que $\sigma \neq 10.0$ años.

b) Se sospecha que la varianza de la distribución de las distancias en kilómetros que un modelo nuevo de automóvil equipado con un motor diesel recorre con 5 litros de combustible es menor que la varianza de la distribución de distancias que recorre el mismo modelo equipado con un motor de gasolina de 6 cilindros, la cual se sabe es $\sigma^2 = 6.25$. Si 72 recorridos de prueba con el modelo diesel tienen una varianza de 4.41, ¿podemos concluir, a un nivel de significancia de 0.05, que la varianza de las distancias recorridas por el modelo que funciona con diesel es menor que la del modelo que funciona con gasolina?

10.73 Se realiza un estudio para comparar el tiempo que les toma a hombres y mujeres ensamblar cierto producto. La experiencia indica que la distribución del tiempo tanto para hombres como para mujeres es aproximadamente normal, pero que la varianza del tiempo para las mujeres es menor que para los hombres. Una muestra aleatoria de los tiempos de 11 hombres y 14 mujeres produce los siguientes datos:

Hombres	**Mujeres**
$n_1 = 11$	$n_2 = 14$
$s_1 = 6.1$	$s_2 = 5.3$

Pruebe la hipótesis de que $\sigma_1^2 = \sigma_2^2$ contra la alternativa de que $\sigma_1^2 > \sigma_2^2$. Utilice un valor P en su conclusión.

10.74 En el ejercicio 10.41 de la página 358 pruebe la hipótesis a un nivel de significancia de 0.05 de que $\sigma_1^2 = \sigma_2^2$ contra la alternativa de que $\sigma_1^2 \neq \sigma_2^2$, donde σ_1^2 y σ_2^2 son las varianzas para el número de organismos por metro cuadrado de agua en los dos lugares diferentes de Cedar Run.

10.75 Remítase al ejercicio 10.39 de la página 358 y pruebe la hipótesis de que $\sigma_1^2 = \sigma_2^2$ contra la alternativa de que $\sigma_1^2 \neq \sigma_2^2$, donde σ_1^2 y σ_2^2 son las varianzas para la duración de las películas producidas por la empresa 1 y la empresa 2, respectivamente. Utilice un valor P.

10.76 Se comparan dos tipos de instrumentos para medir la cantidad de monóxido de azufre en la atmósfera en un experimento sobre la contaminación del aire. Los investigadores desean determinar si los dos tipos de instrumentos proporcionan mediciones con la misma variabilidad. Se registran las siguientes lecturas para los dos instrumentos:

Monóxido de azufre

Instrumento *A*	**Instrumento *B***
0.86	0.87
0.82	0.74
0.75	0.63
0.61	0.55
0.89	0.76
0.64	0.70
0.81	0.69
0.68	0.57
0.65	0.53

Suponga que las poblaciones de mediciones se distribuyen de forma aproximadamente normal y pruebe la hipótesis de que $\sigma_A = \sigma_B$, contra la alternativa de que $\sigma_A \neq \sigma_B$. Use un valor P.

10.77 Se lleva a cabo un experimento para comparar el contenido de alcohol en una salsa de soya en dos líneas de producción diferentes. La producción se supervisa ocho veces al día. A continuación se presentan los datos.

```
Línea de producción 1.
  0.48 0.39 0.42 0.52 0.40 0.48 0.52 0.52
Línea de producción 2.
  0.38 0.37 0.39 0.41 0.38 0.39 0.40 0.39
```

Suponga que ambas poblaciones son normales. Se sospecha que la línea de producción 1 no está produciendo tan consistentemente como la línea 2 en términos de contenido de alcohol. Pruebe la hipótesis de que $\sigma_1 = \sigma_2$ contra la alternativa de que $\sigma_1 \neq \sigma_2$. Utilice un valor P.

10.78 Se sabe que las emisiones de hidrocarburos de los automóviles disminuyeron de forma drástica durante la década de 1980. Se realizó un estudio para comparar las emisiones de hidrocarburos a velocidad estacionaria, en partes por millón (ppm), para automóviles de 1980 y 1990. Se seleccionaron al azar 20 automóviles de cada modelo y se registraron sus niveles de emisión de hidrocarburos. Los datos son los siguientes:

```
Modelos 1980:
141 359 247 940 882 494 306 210 105 880
200 223 188 940 241 190 300 435 241 380
Modelos 1990:
140 160 20 20 223 60 20 95 360 70
220 400 217 58 235 380 200 175 85 65
```

Pruebe la hipótesis de que $\sigma_1 = \sigma_2$ contra la alternativa de que $\sigma_1 \neq \sigma_2$. Suponga que ambas poblaciones son normales. Utilice un valor P.

10.11 Prueba de la bondad de ajuste

A lo largo de este capítulo nos ocupamos de la prueba de hipótesis estadística acerca de parámetros de una sola población, como μ, σ^2 y p. Ahora consideraremos una prueba para determinar si una población tiene una distribución teórica específica. La prueba se basa en el nivel de ajuste que existe entre la frecuencia de ocurrencia de las observaciones en una muestra observada y las frecuencias esperadas que se obtienen a partir de la distribución hipotética.

Para ilustrar lo anterior considere el lanzamiento de un dado. Suponemos que se trata de un dado legal, lo cual equivale a probar la hipótesis de que la distribución de resultados es la distribución uniforme discreta

$$f(x) = \frac{1}{6}, \qquad x = 1, 2, \ldots, 6.$$

Suponga que el dado se lanza 120 veces y que se registra cada resultado. Teóricamente, si el dado está balanceado, esperaríamos que cada cara ocurriera 20 veces. Los resultados se presentan en la tabla 10.4.

Tabla 10.4: Frecuencias observadas y esperadas de 120 lanzamientos de un dado

Cara	**1**	**2**	**3**	**4**	**5**	**6**
Observadas	20	22	17	18	19	24
Esperadas	20	20	20	20	20	20

Al comparar las frecuencias observadas con las frecuencias esperadas correspondientes debemos decidir si es posible que tales discrepancias ocurran como resultado de fluctuaciones del muestreo, de que el dado está balanceado o no es legal o de que la distribución de resultados no es uniforme. Es práctica común referirse a cada resultado posible de un experimento como una celda. En nuestro caso tenemos 6 celdas. A continuación se define el estadístico adecuado en el cual basamos nuestro criterio de decisión para un experimento que incluye k celdas.

Una **prueba de la bondad de ajuste** entre las frecuencias observadas y esperadas se basa en la cantidad.

Prueba de la bondad de ajuste

$$\chi^2 = \sum_{i=1}^{k} \frac{(o_i - e_i)^2}{e_i},$$

donde χ^2 es un valor de una variable aleatoria cuya distribución muestral se aproxima muy de cerca a la distribución chi cuadrada con $v = k - 1$ grados de libertad. Los símbolos o_i y e_i representan las frecuencias observada y esperada, respectivamente, para la i-ésima celda.

El número de grados de libertad asociado con la distribución chi cuadrada que se utiliza aquí es igual a $k - 1$, pues sólo hay $k - 1$ frecuencias de celdas libremente determinadas. Es decir, una vez que se determinan las frecuencias de $k - 1$ celdas, también se determina la frecuencia para la k-ésima celda.

Si las frecuencias observadas se acercan a las frecuencias esperadas correspondientes, el valor χ^2 será pequeño, lo cual indica un buen ajuste. Si las frecuencias observadas difieren de manera considerable de las frecuencias esperadas, el valor χ^2 será grande y el ajuste deficiente. Un buen ajuste conduce a la aceptación de H_0, mientras que un ajuste deficiente conduce a su rechazo. Por lo tanto, la región crítica caerá en la cola derecha de la distribución chi cuadrada. Para un nivel de significancia igual a α encontramos el valor crítico χ^2_α de la tabla A.5 y, entonces, $\chi^2 > \chi^2_\alpha$ constituye la región crítica. **El criterio de decisión que aquí se describe no se debería utilizar a menos que cada una de las frecuencias esperadas sea por lo menos igual a 5**. Esta restricción podría requerir la combinación de celdas adyacentes, lo que dará como resultado una reducción en el número de grados de libertad.

En la tabla 10.4 encontramos que el valor χ^2 es

$$\chi^2 = \frac{(20-20)^2}{20} + \frac{(22-20)^2}{20} + \frac{(17-20)^2}{20} + \frac{(18-20)^2}{20} + \frac{(19-20)^2}{20} + \frac{(24-20)^2}{20} = 1.7.$$

Si usamos la tabla A.5, encontramos $\chi^2_{0.05} = 11.070$ para $v = 5$ grados de libertad. Como 1.7 es menor que el valor crítico, no se rechaza H_0. Concluimos que no hay suficiente evidencia de que el dado está desbalanceado.

Como un segundo ejemplo probemos la hipótesis de que la distribución de frecuencias de la duración de baterías presentadas en la tabla 1.7 de la página 23 se puede aproximar mediante una distribución normal con media $\mu = 3.5$ y desviación estándar $\sigma = 0.7$. Las frecuencias esperadas para las 7 clases (celdas) que se listan en la tabla 10.5 se obtienen calculando las áreas bajo la curva normal hipotética que caen entre los diversos límites de clase.

Tabla 10.5: Frecuencias observadas y esperadas para la duración de las baterías suponiendo normalidad

Límites de clase	o_i		e_i	
1.45 − 1.95	2	7	0.5	8.5
1.95 − 2.45	1		2.1	
2.45 − 2.95	4		5.9	
2.95 − 3.45	15		10.3	
3.45 − 3.95	10		10.7	
3.95 − 4.45	5	8	7.0	10.5
4.45 − 4.95	3		3.5	

Por ejemplo, los valores z que corresponden a los límites de la cuarta clase son

$$z_1 = \frac{2.95 - 3.5}{0.7} = -0.79 \quad \text{y} \quad z_2 = \frac{3.45 - 3.5}{0.7} = -0.07.$$

En la tabla A.3 encontramos que el área entre $z_1 = -0.79$ y $z_2 = -0.07$ es

$$\begin{aligned} \text{área} &= P(-0.79 < Z < -0.07) = P(Z < -0.07) - P(Z < -0.79) \\ &= 0.4721 - 0.2148 = 0.2573. \end{aligned}$$

Por lo tanto, la frecuencia esperada para la cuarta clase es

$$e_4 = (0.2573)(40) = 10.3.$$

Se acostumbra redondear estas frecuencias a un decimal.

La frecuencia esperada para el primer intervalo de clase se obtiene utilizando el área total bajo la curva normal a la izquierda del límite 1.95. Para el último intervalo de clase usamos el área total a la derecha del límite 4.45. Todas las demás frecuencias esperadas se determinan utilizando el método que se describe para la cuarta clase. Observe que combinamos clases adyacentes en la tabla 10.5 donde las frecuencias esperadas son menores que 5 (una regla general en la prueba de la bondad de ajuste). En consecuencia, el número total de intervalos se reduce de 7 a 4, lo cual da como resultado $v = 3$ grados de libertad. Entonces, el valor χ^2 es dado por

$$\chi^2 = \frac{(7-8.5)^2}{8.5} + \frac{(15-10.3)^2}{10.3} + \frac{(10-10.7)^2}{10.7} + \frac{(8-10.5)^2}{10.5} = 3.05.$$

Como el valor χ^2 calculado es menor que $\chi^2_{0.05} = 7.815$ para 3 grados de libertad, no tenemos razón para rechazar la hipótesis nula y concluimos que la distribución normal con $\mu = 3.5$ y $\sigma = 0.7$ proporciona un buen ajuste para la distribución de la duración de las baterías.

La prueba de bondad de ajuste chi cuadrada es un recurso importante, en particular debido a que muchos procedimientos estadísticos en la práctica dependen, en un sentido teórico, de la suposición de que los datos reunidos provienen de un tipo de distribución específico. Como ya se expuso, la suposición de normalidad se hace muy a menudo. En los siguientes capítulos continuaremos haciendo suposiciones de normalidad con el fin de proporcionar una base teórica para ciertas pruebas e intervalos de confianza.

En la literatura hay pruebas para evaluar la normalidad que son más poderosas que la prueba chi cuadrada. Una de tales pruebas es la **prueba de Geary**, la cual se basa en un estadístico muy sencillo que es el cociente de dos estimadores de la desviación estándar de la población σ. Suponga que se toma una muestra aleatoria $X_1, X_2, \ldots, X_n$ de una distribución normal, $N(\mu, \sigma)$. Considere el cociente

$$U = \frac{\sqrt{\pi/2}\sum_{i=1}^{n} |X_i - \bar{X}|/n}{\sqrt{\sum_{i=1}^{n}(X_i - \bar{X})^2/n}}.$$

El lector debería reconocer que el denominador es un estimador razonable de σ sin importar si la distribución es normal o no. El numerador es un buen estimador de σ si la distribución es normal, pero podría sobrestimar o subestimar a σ cuando haya desviaciones de la normalidad. Así, los valores de U que difieren considerablemente de 1.0 representan la señal de que se debe rechazar la hipótesis de normalidad.

Para muestras grandes una prueba razonable se basa en la normalidad aproximada de U. El estadístico de prueba es, entonces, una estandarización de U dada por

$$Z = \frac{U-1}{0.2661/\sqrt{n}}.$$

Desde luego, el procedimiento de prueba incluye la región crítica bilateral. Calculamos un valor de z a partir de los datos y no rechazamos la hipótesis de normalidad cuando

$$-z_{\alpha/2} < Z < z_{\alpha/2}.$$

En la bibliografía se cita un artículo que trata sobre la prueba de Geary (Geary, 1947).

10.12 Prueba de independencia (datos categóricos)

El procedimiento de prueba de chi cuadrada que se presentó en la sección 10.11 también se puede usar para probar la hipótesis de independencia de dos variables de clasificación. Suponga que deseamos determinar si las opiniones de los votantes residentes del estado de Illinois respecto a una nueva reforma fiscal son independientes de sus niveles de ingreso. Los sujetos de una muestra aleatoria de 1000 votantes registrados del estado de Illinois se clasifican de acuerdo con su posición en las categorías de ingreso bajo, medio o alto, y si están a favor o no de la nueva reforma fiscal. Las frecuencias observadas se presentan en la tabla 10.6, la cual se conoce como **tabla de contingencia**.

Tabla 10.6: Tabla de contingencia 2×3

	Nivel de ingreso			
Reforma fiscal	**Bajo**	**Medio**	**Alto**	**Total**
A favor	182	213	203	598
En contra	154	138	110	402
Total	336	351	313	1000

Una tabla de contingencia con r renglones y c columnas se denomina tabla $r \times c$ ("$r \times c$" se lee "r por c"). Los totales de renglones y columnas en la tabla 10.6 se denominan **frecuencias marginales**. Nuestra decisión de aceptar o rechazar la hipótesis nula, H_0, de que la opinión de un votante respecto a la nueva reforma fiscal es independiente de su nivel de ingreso, se basa en qué tan bien se ajusten las frecuencias observadas en cada una de las 6 celdas de la tabla 10.6 y en las frecuencias que esperaríamos para cada celda si supusiéramos que H_0 es verdadera. Para encontrar estas frecuencias esperadas definamos los siguientes eventos:

L: Una persona seleccionada está en el nivel de ingresos bajo.

M: Una persona seleccionada está en el nivel de ingresos medio.

H: Una persona seleccionada está en el nivel de ingresos alto.

F: Una persona seleccionada está a favor de la nueva reforma fiscal.

A: Una persona seleccionada está en contra de la nueva reforma fiscal.

Podemos usar las frecuencias marginales para listar las siguientes estimaciones de probabilidad:

$$P(L) = \frac{336}{1000}, \qquad P(M) = \frac{351}{1000}, \qquad P(H) = \frac{313}{1000},$$
$$P(F) = \frac{598}{1000}, \qquad P(A) = \frac{402}{1000}.$$

Ahora bien, si H_0 es verdadera y las dos variables son independientes, deberíamos tener

$$P(L \cap F) = P(L)\,P(F) = \left(\frac{336}{1000}\right)\left(\frac{598}{1000}\right),$$
$$P(L \cap A) = P(L)\,P(A) = \left(\frac{336}{1000}\right)\left(\frac{402}{1000}\right),$$

$$P(M \cap F) = P(M)\,P(F) = \left(\frac{351}{1000}\right)\left(\frac{598}{1000}\right),$$
$$P(M \cap A) = P(M)\,P(A) = \left(\frac{351}{1000}\right)\left(\frac{402}{1000}\right),$$
$$P(H \cap F) = P(H)\,P(F) = \left(\frac{313}{1000}\right)\left(\frac{598}{1000}\right),$$
$$P(H \cap A) = P(H)\,P(A) = \left(\frac{313}{1000}\right)\left(\frac{402}{1000}\right).$$

Las frecuencias esperadas se obtienen multiplicando la probabilidad de cada celda por el número total de observaciones. Como antes, redondeamos estas frecuencias a un decimal. Así, se estima que el número esperado de votantes de bajo ingreso en nuestra muestra que favorecen la reforma fiscal es

$$\left(\frac{336}{1000}\right)\left(\frac{598}{1000}\right)(1000) = \frac{(336)(598)}{1000} = 200.9$$

cuando H_0 es verdadera. La regla general para obtener la frecuencia esperada de cualquier celda es dada por la siguiente fórmula:

$$\text{frecuencia esperada} = \frac{(\text{total por columna}) \times (\text{total por renglón})}{\text{gran total}}$$

En la tabla 10.7 la frecuencia esperada para cada celda se registra entre paréntesis, a un lado del valor observado verdadero. Observe que las frecuencias esperadas en cualquier renglón o columna se suman al total marginal apropiado. En nuestro ejemplo necesitamos calcular sólo las dos frecuencias esperadas en el renglón superior de la tabla 10.7 y luego calcular las otras mediante sustracción. El número de grados de libertad asociados con la prueba chi cuadrada que aquí se usa es igual al número de frecuencias de celdas que se pueden llenar libremente cuando se nos proporcionan los totales marginales y el gran total, y en este caso ese número es 2. Una fórmula sencilla que proporciona el número correcto de grados de libertad es

$$v = (r - 1)(c - 1).$$

Tabla 10.7: Frecuencias observadas y esperadas

	Nivel de ingreso			
Reforma fiscal	**Bajo**	**Medio**	**Alto**	**Total**
A favor	182 (200.9)	213 (209.9)	203 (187.2)	598
En contra	154 (135.1)	138 (141.1)	110 (125.8)	402
Total	336	351	313	1000

Por lo tanto, para nuestro ejemplo $v = (2 - 1)(3 - 1) = 2$ grados de libertad. Para probar la hipótesis nula de independencia usamos el siguiente criterio de decisión:

Prueba de independencia

Calcule

$$\chi^2 = \sum_i \frac{(o_i - e_i)^2}{e_i},$$

donde la sumatoria se extiende a todas las celdas rc en la tabla de contingencia $r \times c$.

Si $\chi^2 > \chi^2_\alpha$ con $v = (r - 1)(c - 1)$ grados de libertad, rechace la hipótesis nula de independencia al nivel de significancia α; en otro caso no la rechace.

Al aplicar este criterio a nuestro ejemplo encontramos que

$$\begin{aligned}\chi^2 &= \frac{(182 - 200.9)^2}{200.9} + \frac{(213 - 209.9)^2}{209.9} + \frac{(203 - 187.2)^2}{187.2} \\ &\quad + \frac{(154 - 135.1)^2}{135.1} + \frac{(138 - 141.1)^2}{141.1} + \frac{(110 - 125.8)^2}{125.8} = 7.85, \\ P &\approx 0.02.\end{aligned}$$

En la tabla A.5 encontramos que $\chi^2_{0.05} = 5.991$ para $v = (2 - 1)(3 - 1) = 2$ grados de libertad. Rechazamos la hipótesis nula y concluimos que la opinión de un votante respecto a la reforma fiscal y su nivel de ingresos no son independientes.

Es importante recordar que el estadístico sobre el cual basamos nuestra decisión tiene una distribución que sólo se aproxima por la distribución chi cuadrada. Los valores χ^2 calculados dependen de las frecuencias de las celdas y, en consecuencia, son discretos. La distribución chi cuadrada continua parece aproximarse muy bien a la distribución de muestreo discreta de χ^2, siempre y cuando el número de grados de libertad sea mayor que 1. En una tabla de contingencia de 2×2, donde sólo tenemos 1 grado de libertad, se aplica una corrección llamada **corrección de Yates para continuidad**.

La fórmula corregida entonces se convierte en

$$\chi^2(\text{corregida}) = \sum_i \frac{(|o_i - e_i| - 0.5)^2}{e_i}.$$

Si las frecuencias de las celdas esperadas son grandes, los resultados corregidos y sin corrección son casi iguales. Cuando las frecuencias esperadas están entre 5 y 10, se debe aplicar la corrección de Yates. Para frecuencias esperadas menores que 5 se debería utilizar la prueba exacta de Fisher-Irwin. Un análisis de esta prueba se puede encontrar en *Basic Concepts of Probability and Statistics* de Hodges y Lehmann (2005; véase la bibliografía). Sin embargo, la prueba de Fisher-Irwin se puede evitar seleccionando una muestra grande.

10.13 Prueba de homogeneidad

Cuando probamos la independencia en la sección 10.12 seleccionamos una muestra aleatoria de 1000 votantes, y determinamos al azar los totales de renglón y de columna para nuestra tabla de contingencia. Otro tipo de problema para el que se aplica el método de la sección 10.12 es aquel en el cual los totales de renglón y de columna están predeterminados. Suponga, por ejemplo, que decidimos de antemano seleccionar 200 demócratas, 150 republicanos y 150 independientes entre los votantes del estado de Carolina del Norte y registrar si están a favor de una iniciativa de ley para el aborto, si están en contra o si están indecisos. Las respuestas observadas se incluyen en la tabla 10.8.

Tabla 10.8: Frecuencias observadas

	Afiliación política			
Ley para el aborto	**Demócrata**	**Republicano**	**Independiente**	**Total**
A favor	82	70	62	214
En contra	93	62	67	222
Indeciso	25	18	21	64
Total	200	150	150	500

Ahora bien, en vez de hacer una prueba de independencia, probamos la hipótesis de que las proporciones de población dentro de cada renglón son iguales. Es decir, probamos la hipótesis de que las proporciones de demócratas, republicanos e independientes que están a favor de la ley para el aborto son iguales; las proporciones de cada afiliación política contra la ley son iguales y las proporciones de cada afiliación política que están indecisos son iguales. Básicamente nos interesamos en determinar si las tres categorías de votantes son **homogéneas** en lo que se refiere a sus opiniones acerca de la iniciativa de ley para el aborto. A esta prueba se le conoce como prueba de homogeneidad.

Al suponer homogeneidad de nuevo calculamos las frecuencias esperadas de las celdas multiplicando los totales de renglón y de columna correspondientes y después dividiendo entre el gran total. Luego continuamos el análisis utilizando el mismo estadístico chi cuadrada como antes. Ilustramos este proceso en el siguiente ejemplo para los datos de la tabla 10.8.

Ejemplo 10.14: Con respecto a los datos de la tabla 10.8 pruebe la hipótesis de que las opiniones en cuanto a la propuesta de ley para el aborto son las mismas en cada afiliación política. Utilice un nivel de significancia de 0.05.

***Solución*:**

1. H_0: Para cada opinión las proporciones de demócratas, republicanos e independientes son iguales.
2. H_1: Para al menos una opinión las proporciones de demócratas, republicanos e independientes no son iguales.
3. $\alpha = 0.05$.
4. Región crítica: $\chi^2 > 9.488$ con $v = 4$ grados de libertad.
5. Cálculos: necesitamos calcular las 4 frecuencias de las celdas usando la fórmula de las frecuencias de las celdas esperadas de la página 375. Todas las demás frecuencias se obtienen mediante sustracción. Las frecuencias de las celdas observadas y esperadas se muestran en la tabla 10.9.

Tabla 10.9: Frecuencias observadas y esperadas

	Afiliación política			
Ley para el aborto	**Demócrata**	**Republicano**	**Independiente**	**Total**
A favor	82 (85.6)	70 (64.2)	62 (64.2)	214
En contra	93 (88.8)	62 (66.6)	67 (66.6)	222
Indeciso	25 (25.6)	18 (19.2)	21 (19.2)	64
Total	200	150	150	500

Así,

$$\chi^2 = \frac{(82-85.6)^2}{85.6} + \frac{(70-64.2)^2}{64.2} + \frac{(62-64.2)^2}{64.2} + \frac{(93-88.8)^2}{88.8} + \frac{(62-66.6)^2}{66.6} + \frac{(67-66.6)^2}{66.6} + \frac{(25-25.6)^2}{25.6} + \frac{(18-19.2)^2}{19.2} + \frac{(21-19.2)^2}{19.2} = 1.53.$$

6. Decisión: No rechazar H_0. No hay suficiente evidencia para concluir que la proporción de demócratas, republicanos e independientes difiere para cada opinión expresada.

Prueba para varias proporciones

El estadístico chi cuadrada para probar la homogeneidad también se puede aplicar cuando se prueba la hipótesis de que k parámetros binomiales tienen el mismo valor. Por lo tanto, se trata de una extensión de la prueba que se presentó en la sección 10.9 para determinar las diferencias entre dos proporciones a una prueba para determinar diferencias entre k proporciones. En consecuencia, nos interesamos en probar la hipótesis nula

$$H_0\colon\ p_1 = p_2 = \cdots = p_k$$

contra la hipótesis alternativa H_1 de que las proporciones de la población *no son todas iguales.* Para ejecutar esta prueba primero observamos muestras aleatorias independientes de tamaños $n_1, n_2, \ldots, n_k$ de las k poblaciones y ordenamos los datos en una tabla de contingencia $2 \times k$, la tabla 10.10.

Tabla 10.10: k muestras binomiales independientes

Muestra:	**1**	**2**	$\cdots$	$\boldsymbol{k}$
Éxitos	x_1	x_2	$\cdots$	x_k
Fracasos	$n_1 - x_1$	$n_2 - x_2$	$\cdots$	$n_k - x_k$

De acuerdo con si los tamaños de las muestras aleatorias fueron predeterminados o si ocurrieron al azar, el procedimiento de prueba es idéntico a la prueba de homogeneidad o a la prueba de independencia. Por lo tanto, las frecuencias de las celdas esperadas se calculan como antes y se sustituyen junto con las frecuencias observadas en el estadístico chi cuadrada

$$\chi^2 = \sum_i \frac{(o_i - e_i)^2}{e_i},$$

con

$$v = (2-1)(k-1) = k-1$$

grados de libertad.

Al seleccionar la región crítica apropiada de la cola superior de la forma $\chi^2 > \chi^2_\alpha$ podemos llegar ahora a una decisión respecto a H_0.