Unidad 1

- Inferencia Estadística con Muestras Grandes

8.1 Resumen
8.2 El objetivo de la estadística, la inferencia
8.3 Tipos de estimadores
8.4 Evaluación de la bondad en un estimador puntual
8.5 Evaluación de un estimador por intervalo
8.6 Estimación puntual de la media de una población
8.7 Estimación por intervalo de la media de una población
8.8 Estimación con muestras grandes
8.9 Estimación de la diferencia de dos medias
8.10 Estimación del parámetro de una población binomial
8.11 Estimación de la diferencia entre 2 proporciones
8.12 Selección del tamaño de muestra
8.13 Prueba estadística de una hipótesis
8.14 Pruebas estadísticas con muestras grandes
8.15 Algunos comentarios sobre la teoría de pruebas de hipótesis
8.16 Resumen

8 Inferencia estadística con muestras grandes

8.1 Resumen

En los siete capítulos anteriores se ha preparado el terreno para el estudio del objetivo de este texto, el desarrollo y el entendimiento de la inferencia estadística y el papel que ésta juega en la toma de decisiones en los negocios. En el capítulo 1 se estableció que la labor del estadístico consiste en hacer inferencias acerca de las observaciones de una población basándose en la información contenida en una muestra. En el capítulo 3 se vió como describir un conjunto de observaciones para bosquejar inferencias acerca de la población. En el capítulo 4, se discutió el concepto de probabilidad, concepto fundamental para el mecanismo de la inferencia. Se continuó con tres capítulos acerca de distribuciones de probabilidad, una discusión general en el capítulo 5, algunas distribuciones de probabilidad discretas en el capítulo 6 y en el capítulo 7, la distribución normal.

Para empezar a pensar en términos de la inferencia estadística, se introdujo en el capítulo 6 una aplicación de la distribución binomial a un problema de muestreo de aceptación y a la prueba de una hipótesis acerca de la eficiencia de una vacuna para el resfriado. Estos temas se ilustraron en ejercicios y ejemplos del capítulo 7. Ahora se hará uso de los cimientos que se han construído para estudiar los conceptos fundamentales de la inferencia estadística.

Quizá la más importante de las contribuciones a la preparación para el estudio de la inferencia estadística ha sido el teorema central del límite presentado en el capítulo 7. Este teorema justifica la normalidad aproximada de la distribución de probabilidad de la media muestral cuando el tamaño de la muestra es grande. Pero aún más importante, en este capítulo será usado para justificar la normalidad aproximada de los estimadores y las variables de decisión que se presentan.

8.2 El objetivo de la estadística, la inferencia

La inferencia, particularmente la toma de decisiones y la predicción, ha jugado un papel muy importante en la vida del hombre desde la antigüedad. Cada persona se encuentra con situaciones que requieren predicciones. Al gobierno le preocupa cuál será el flujo de oro hacia Europa. Al inversionista le gustaría predecir el comportamiento de la bolsa de valores. A algunos industriales les interesa el resultado de un experimento para saber si un nuevo tipo de acero es más resistente a cambios de temperatura. Las amas de casa desean saber si en verdad el detergente A es más efectivo que el detergente B en su lavadora. Estas inferencias se basan supuestamente en información relevante en forma de datos u observaciones.

En muchas situaciones prácticas la información relevante es abundante, aparentemente inconsistente y, en muchos casos, abrumadora. Como consecuencia nuestra decisión o predicción es a menudo poco mejor que una adivinanza. Basta un vistazo a las revistas especializadas en el mercado de valores para observar la diversidad de opiniones de los expertos acerca del comportamiento futuro del precio de las acciones. Asimismo, si los científicos y los ingenieros analizan visualmente un conjunto de datos, se producirán opiniones contradictorias acerca de las conclusiones a que puede llegarse como resultado de un experimento. Aunque mucha gente piensa que su mecanismo natural para hacer inferencias es muy bueno, la experiencia indica que la mayoría de la gente es incapaz de utilizar grandes cantidades de datos, evaluar mentalmente cada porción de información relevante y terminar con una buena inferencia. (Pruebe usted mismo su habilidad para hacer inferencias usando los ejercicios en este capítulo y el siguiente. Analice visualmente los datos y haga su inferencia antes de usar el procedimiento estadístico adecuado. Compare los resultados.) Se concluye que es deseable el estudio de sistemas para hacer inferencia y éste es precisamente el objetivo de los estadísticos matematicos.

Aunque, con toda intención, en capítulos anteriores se han mencionado algunas nociones de inferencia estadística, es conveniente organizar estos conceptos, haciendo una presentación elemental de algunas ideas básicas de la inferencia estadística.

El objetivo de la estadística es hacer inferencias acerca de una población con base en la información contenida en una muestra.

Puesto que las poblaciones se caracterizan por medidas descriptivas numéricas llamadas **parámetros,** la inferencia estadística se ocupa de hacer inferencias acerca de los parámetros de una población. Parámetros típicos de una población son la media, la desviación estándar, el área bajo la distribución de probabilidad a partir de un valor de una variable aleatoria, o el área entre dos valores de la variable. Es posible reformular todos los problemas mencionados en el primer párrafo de esta sección en términos de una población con parámetros de interes específicos.

Los métodos para hacer inferencias acerca de los parámetros pueden clasificarse en una de dos categorías. Pueden tomarse **decisiones** acerca del valor del parámetro como en el ejemplo de muestreo de aceptación y prueba de hipótesis descrito en el capítulo 6, o bien se puede predecir o **estimar** el valor del parámetro. Aunque algunos estadísticos contemplan la estimación como un problema de toma de decisión, es conveniente conservar las dos categorías y en particular concentrarse separadamente en estimación y pruebas de hipótesis.

No sería adecuado hablar de los objetivos y de los tipos de inferencia estadística sin referirse a alguna medida de la bondad de los procedimientos de inferencia. Es posible definir numerosos métodos objetivos para hacer inferencias además de los procedimientos propios basados en la intuición. Por lo tanto se debe tener una medida para poder comparar la bondad de un estimador con la de otro. Más aún, se desea establecer la bondad de una inferencia en una situación real dada. Predecir que el precio de un bono financiero será de $80 el próximo lunes es insuficiente y pocos se verán motivados a tomar la acción de comprar o vender. Es deseable saber también si el estimador es correcto dentro de más o menos $1, $2 ó $10.

La inferencia estadística en una situación práctica contempla dos aspectos: (1) la inferencia, y (2) una medida de su bondad.

Cuál de los métodos de inferencia debe usarse, es decir, ¿se requiere estimar el parámetro? O ¿debe probarse una hipótesis acerca de su valor? La respuesta está determinada por la situación práctica a considerar y en ocasiones es cuestión de preferencia personal. Algunas personas prefieren probar teorías acerca de los parámetros y otras prefieren expresar sus inferencias en forma de un estimador. Ambas, estimación y prueba de hipótesis, se usan frecuentemente en la literatura científica. Por lo tanto, ambas deben ser incluídas en esta discusión.

8.3 Tipos de estimadores

Los procedimientos de estimación pueden ser divididos en dos tipos, estimación puntual y estimación por intervalo. Suponga que como gerente de un supermercado está usted preocupado por la desaparición de mercancía a causa de robo. Para estimar la tasa de desaparición se toma una muestra de n artículos y se calcula para cada artículo

$$\text{tasa de desaparición} = \frac{\text{inventario} - \text{existencia}}{\text{inventario}}$$

$$= \frac{\text{pérdida}}{\text{inventario}}$$

A partir de estos datos muestrales es posible obtener dos tipos de estimadores de la tasa de desaparición media μ para todos los artículos en la tienda. Si la media muestral es .20 este número puede usarse como un estimador de μ. Es

decir que la estimación de la pérdida media es de un 20% del inventario. O puede estimarse que la pérdida media está en el intervalo de .15 a .25 o, equivalentemente, se estima que la pérdida media está entre 15% y 25% del inventario. El primer tipo de estimación se llama **estimación puntual** puesto que ese solo número que representa la estimación puede asociarse con un punto en una recta. El segundo tipo, con base en dos puntos que definen un intervalo en la recta, es llamado **estimación por intervalo.** Cada uno de estos métodos de estimación serán considerados aquí.

Un procedimiento de estimación puntual utiliza la información en una muestra y la sintetiza en un solo número o punto que estima el parámetro de la población de interés. La estimación se lleva a cabo por medio de un **estimador.** Un estimador es una regla que indica como calcular la estimación en base a la información de una muestra. Generalmente se expresa por medio de una fórmula. Por ejemplo, la media muestral

$$\bar{y} = \frac{\sum_{i=1}^{n} y_i}{n}$$

es un estimador de la media de la población μ e indica exactamente como obtener el valor numérico de la estimación una vez que se conocen los valores de la muestra $y_1, y_2, \ldots, y_n$. En cambio un estimador por intervalo usa los datos muestrales para calcular dos puntos entre los cuales se espera que se encuentre el valor del parámetro de la población que se quiere estimar.

Definición

Un **estimador puntual** del parámetro de una población es una regla que indica como calcular un número con base en los datos muestrales. Al número resultante se le llama **estimación puntual** del parámetro.

Definición

Un **estimador por intervalo** es una regla que indica como calcular dos números con base en los datos muestrales.

Definición

Cuando se usa un estimador por intervalo para estimar el parámetro de una población, el par de números que se obtiene se llama **estimación por intervalo** o **intervalo de confianza.** El número mayor, que indica el extremo superior del intervalo se denomina **límite superior de confianza (LSC).** Similarmente el número extremo inferior del intervalo se denomina **límite inferior de confianza (LIC).**

Ambos tipos de estimaciones se usan comúnmente para encuestas de preferencia del consumidor pero para este caso las estimaciones puntuales son más usadas. Cuando la estimación surge en el contexto de experimentación industrial resulta más común usar intervalos de confianza.

8.4 Evaluación de la bondad en un estimador puntual

Recuerde el ejemplo de la desaparición de artículos por robo en un supermercado. Supóngase que la media muestral resulta ser .20 y que usted reporta este valor a la dirección del supermercado. ¿Quedará el director satisfecho con este valor? o preguntará —¿más o menos cuánto? En otras palabras, ¿qué tanta fé tendrá el director en esta estimación? ¿Cuál es la precisión de este estimador de μ? ¿Qué tan alejado de μ podría estar este estimador?

Para responder considérese una analogía que puede ayudar a explicar el razonamiento empleado para evaluar la bondad de un estimador puntual. La estimación puntual puede compararse, en muchos aspectos, a disparar un revólver hacia un blanco. El estimador, que produce estimaciones, es similar al revólver. Una particular estimación es similar a la bala disparada; y el parámetro de interés es similar al centro de la diana de tiro. Tomar una muestra de la población y estimar el valor del parámetro es equivalente a disparar un balazo hacia la diana. Suponga que un tirador dispara un balazo y atraviesa el centro exacto de la diana de tiro. ¿Podría considerarse que se trata de un excelente tirador? La respuesta es no, ya que muy pocas personas consentirían en sostener el blanco para un segundo disparo. En cámbio si un millón de disparos sucesivos atravesaran el centro mismo de la diana, se tendría suficiente confianza en el tirador como para sostener el blanco para el siguiente disparo si se ofreciera una recompensa adecuada. El hecho que se desea subrayar aquí es que no es posible evaluar un procedimiento de estimación con base en una sola estimación. Por el contrario, deben observarse los resultados de usar el procedimiento de estimación repetidas veces al estimar el mismo parámetro de la misma población. En la analogía del tirador sería como observar qué tan cerca se distribuyen los balazos alrededor del centro del blanco. De hecho como las estimaciones son números, para evaluar la bondad del estimador se construiría la distribución de frecuencias de las estimaciones obtenidas al tomar repetidas muestras y calcular el valor del estimador, y se juzgaría de qué manera esta distribución de frecuencias se centra en el parámetro de interés.

Como ilustración considere el experimento del lanzamiento de un dado que se presentó en el capítulo 7, en donde se generaron 100 muestras de tamaño $n = 5$ cada una y se calculó la media para cada muestra. Como en este caso se conoce el valor medio μ del número que aparece en la cara superior del dado ($\mu = 3.5$), es posible usar los resultados de aquel experimento para ver qué tan bien se estima con base en una muestra de tamaño $n = 5$.

El histograma de frecuencias de las 100 medias muestrales se muestra en la figura 8.1. Observe como las estimaciones se agrupan alrededor de la media de la población $\mu = 3.5$ y van de 1.3 a 5.5. Es claro como esta distribución de estimaciones dice algo acerca de qué tan buena sería una estimación de μ si se tomara una muestra de tamaño $n = 5$ observaciones y se calculara la media muestral $\bar{y}$.

Si para cada estimador se tuviera un histograma de frecuencias relativas (como el de la figura 8.1) que exhibiera el comportamiento del estimador para

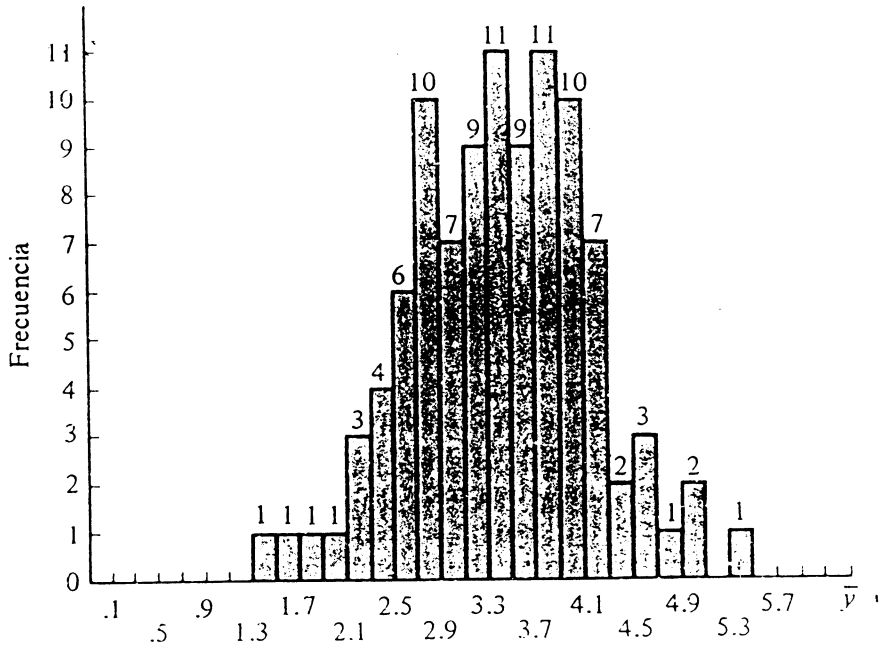


Figura 8.1 *Histograma de medios muestrales del experimento de lanzamiento de un dado en la sección 7.2*

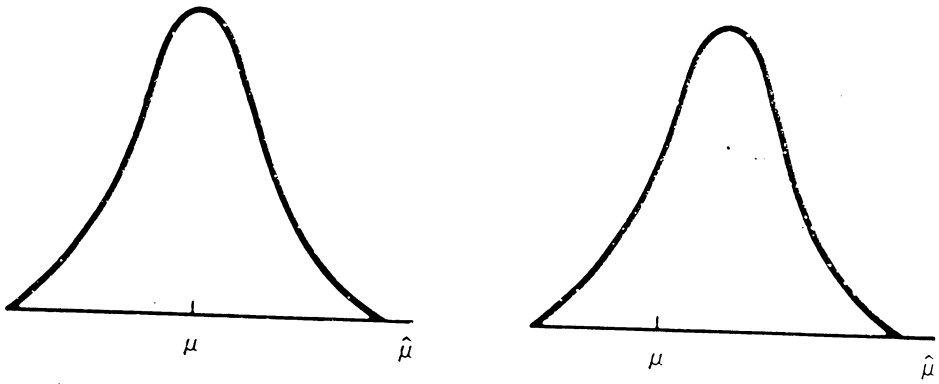
muestras sucesivas del mismo tamaño, se conocerían entonces las características del estimador. Si en lugar de tomar 100 muestras y calcular 100 valores de $\bar{y}$, como en el estudio ilustrado en la figura 8.1, se pensase en tomar miles de muestras, tantas como se desee. Si fuera posible tomar un número infinito de muestras de tamaño fijo, la distribución de frecuencias relativas de las $\bar{y}$'s sería la distribución de probabilidad, o, como se le llama frecuentemente, la **distribución muestral** del estimador $\bar{y}$.

Definición

A la distribución de probabilidad de un estimador se le llama la **distribución muestral** del estimador.

La obtención de las distribuciones muestrales de varios estimadores es el trabajo de los investigadores en estadística. Afortunadamente, la distribución muestral de la mayoría de los estimadores puntuales usuales es conocida.

Ahora que se tiene que las propiedades de un estimador puntual se encuentran en su distribución de probabilidad, surge la pregunta acerca de cuáles propiedades son las más deseables en un estimador. En esencia, son dos las características principales de un estimador. Primero se desea que la distribución de las estimaciones se centre alrededor del parámetro de interés. Por ejemplo, si se está estimando μ, se desearía que la distribución muestral del estimador estuviera centrada en μ, como se ilustra en la figura 8.2(a). A tal estimador se le denomina **insesgado.** La distribución muestral de un estimador **sesgado** se ilustra en la figura 8.2(b). (*Nota:* Se usa una tilde (una testilla) sobre el parámetro para simbolizar un estimador del parámetro.)



(a) Estimador insesgado (b) Estimador sesgado

Figura 8.2 *Distribuciones de estimadores insesgados y sesgados*

Definición

El estimador del parámetro de una población es **insesgado** si la media de su distribución muestral es igual al parámetro. En otro caso se dice que el estimador es **sesgado.**

La segunda propiedad deseable en un estimador puntual es que la desviación estándar de su distribución muestral sea pequeña. Entonces se desea que la dispersión de las estimaciones sea tan pequeña como sea posible. Para la mayoría de los estimadores la desviación estándar de la distribución muestral es controlable. Esto es, es posible hacer la desviación estándar (que mide la dispersión de las estimaciones) tan pequeña como se desee al aumentar el tamaño de la muestra.

Por ejemplo, en el capítulo 7 se vió que debido al teorema central del límite, la distribución muestral de la media muestral $\bar{y}$ se distribuye aproximadamente normal con media μ y desviación estándar $\sigma/\sqrt{n}$ (μ y σ son la media y la desviación estándar de la población de la cual se tomó la muestra y n es el tamaño de la muestra). En consecuencia es posible hacer a la desviación estándar de la distribución muestral de $\bar{y}$, que se denotará por $\sigma_{\bar{y}}$, tan pequeña como se desee al aumentar el tamaño de la muestra n. En la figura 8.3 se ilustra como la distribución muestral de $\bar{y}$ depende del tamaño de la muestra n. En la figura 8.3 se exhiben 3 distribuciones muestrales de $\bar{y}$ dibujadas en la misma gráfica. Las distribuciones muestrales, correspondientes a los tamaños de muestra $n = 5$, $n = 20$ y $n = 80$ muestran como la dispersión de las distribuciones muestrales decrece conforme n aumenta. También se ilustra el porqué una estimación basada en $n = 80$ observaciones estará, con una alta probabilidad, bastante más cerca de μ que una estimación basada en una muestra de $n = 5$.

En una situación real, es posible que se sepa que la distribución muestral del estimador se centra en el parámetro que se desea estimar, pero no se conoce el valor del parámetro. Lo único con lo que se cuenta es con la estimación calculada a partir de las n observaciones de la muestra. ¿Qué tan lejos estará la particular estimación del parámetro estimado? Puesto que el parámetro usualmente se encuentra en el centro de la distribución muestral (que es usualmente la media de la distribución), la distancia entre la estimación y el pará-

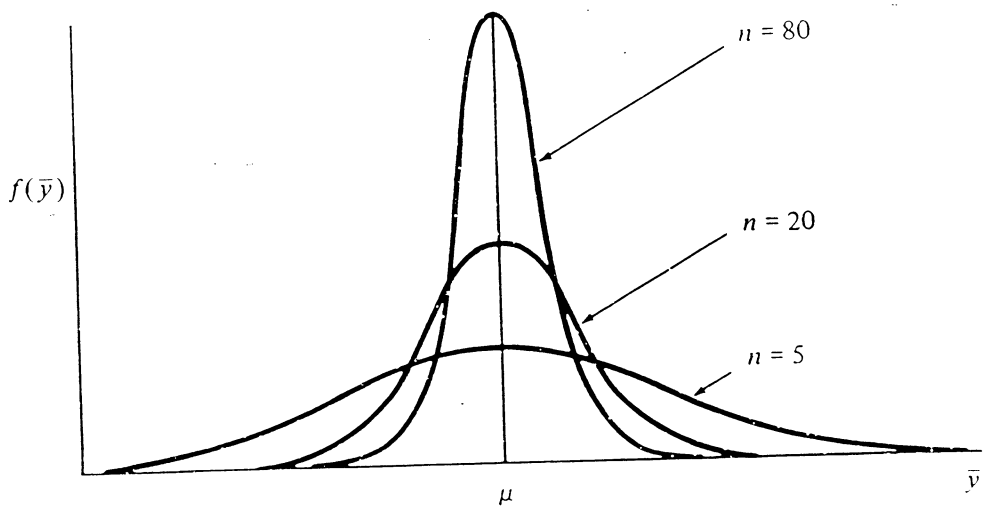


Figura 8.3 *Distribuciones muestrales para* $\bar{y}$; $n = 5$, 20 *y* 80

metro, a la que se le llama **error de estimación,** es menor o igual a la distancia entre el centro y las colas de la distribución.

Definición

La distancia entre una estimación y el valor del parámetro se llama **error de estimación.**

De hecho el error de estimación debe ser menor que dos desviaciones estándar de la distribución muestral, con una probabilidad de al menos .75 (teorema de Tchebysheff), y en una gran cantidad de casos con una probabilidad de .95 (la regla empírica). Por ejemplo, la distribución muestral de la media muestral es aproximadamente normal, con media μ y desviación estándar $\sigma/\sqrt{n}$, (teorema central del límite) y, como se observa en la figura 8.4, la probabilidad de que una estimación caiga dentro de $2\sigma/\sqrt{n}$ hacia abajo o hacia arriba de μ es aproximadamente igual a .95. Equivalentemente la probabilidad de que el error de estimación sea menor que $2\sigma/\sqrt{n}$ es aproximadamente igual a .95.

Para resumir, las propiedades de un estimador puntual están caracterizadas por su distribución muestral. Se preferirán estimadores que sean insesgados y tengan desviación estándar pequeña. Si no es satisfactoria la desviación estándar de un estimador, ésta puede reducirse aumentando el tamaño de muestra. El error de estimación—la distancia entre la estimación y el parámetro estimado—debe ser menor que dos desviaciones estándar de la distribución muestral, con una probabilidad de al menos .75 y en muchos casos cercana a .95.

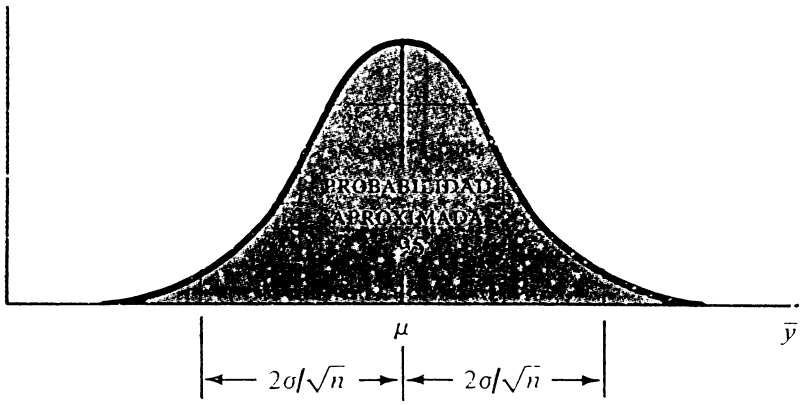


Figura 8.4 *La distribución muestral de* $\bar{y}$

8.5 Evaluación de un estimador por intervalo

El construir una estimación por intervalo es como lanzar un aro para ensartarlo en una estaca. En este caso, el parámetro que se desea estimar corresponde a la estaca y el intervalo al aro que se lanza. Cada vez que se toma una muestra y se construye el intervalo de confianza se espera que el parámetro a estimar quede contenido dentro de éste. Por supuesto esto no ocurrirá para cada muestra. La probabilidad de que un intervalo contenga al parámetro que se estima se denomina **coeficiente de confianza.**

Definición

La probabilidad de que un intervalo de confianza contenga al parámetro que se estima se denomina **coeficiente de confianza.**

Como ejemplo práctico, puede suponerse que se desea estimar la utilidad semanal promedio para una compañía pequeña. Si se tomaran 10 muestras cada una de $n = 20$ observaciones semanales y para cada muestra se construyera un intervalo de confianza para la media μ de la población, es posible que los intervalos que se obtengan ocurran de manera similar a los que se muestran en la figura 8.5. Las líneas horizontales representan los 10 intervalos y la línea vertical representa el verdadero valor de la utilidad semanal media. Nótese que, salvo uno, todos los intervalos contienen a μ, para estas muestras particulares.

Entonces, ¿cuál es un buen intervalo de confianza? La respuesta es aquel que tenga un coeficiente de confianza alto, cercano a uno, y que sea tan estrecho como sea posible. Entre más estrecho sea el intervalo se tendrá localizado el parámetro estimado de manera más precisa. Entre mayor el coeficiente de confianza, más seguridad se tiene de que un intervalo en particular pueda contener al parámetro estimado. Recuérdese que el coeficiente de confianza dá la probabilidad de que la estimación por intervalo produzca límites que con-

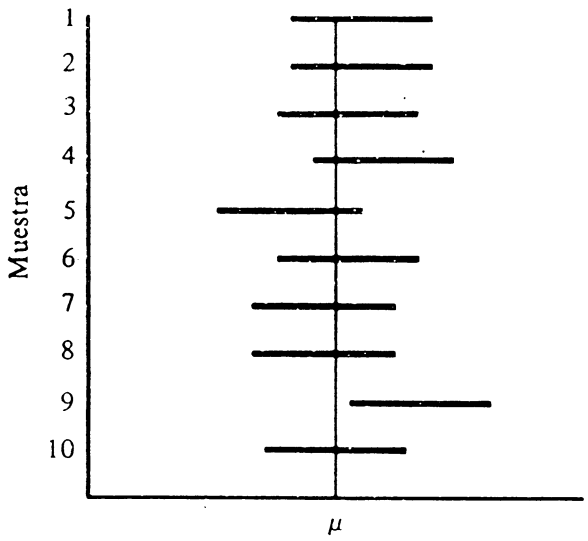


Figura 8.5 *Diez intervalos de confianza para la ganancia semanal media (cada uno basado en una muestra de $n = 20$ observaciones)*

tengan al parámetro estimado. Es decir, proporciona una medida de la confianza que se tiene en los límites de confianza construídos a partir de los datos en la muestra. En resumen la amplitud del intervalo y su coeficiente de confianza asociado miden la bondad de un intervalo de confianza.

¿Cuál es el efecto del tamaño de muestra sobre la amplitud del intervalo? Entre mayor sea la muestra se tiene mayor información para construir el intervalo de confianza. Entonces, para un coeficiente de confianza dado, entre más grande sea la muestra menor será la longitud del intervalo que resulte.

8.6 Estimación puntual de la media de una población

Algunos problemas de decisión en los negocios requieren a menudo de la estimación de la media μ de una población. Por ejemplo, suponga que se está interesado en la producción diaria media de una línea de ensamblado, o en la resistencia promedio de un nuevo tipo de acero, o en el número promedio de accidentes por mes en una fábrica, o bien en la demanda promedio para un nuevo producto. En todos estos casos la estimación de μ resulta una importante aplicación práctica de la inferencia estadística y resulta también una excelente ilustración de los principios de estimación puntual discutidos en la sección 8.4.

Se dispone de varios estimadores para estimar la media μ de una población. Entre estos se tiene la mediana muestral, el promedio entre la máxima y la mínima observación en la muestra y la media muestral $\bar{y}$. Cada uno de éstos tiene asociada una distribución muestral generada por muestreo repetitivo y dependiendo de la población y del problema práctico en cuestión, cada uno

presenta ventajas y desventajas. La mediana muestral y el promedio de las observaciones extremas son frecuentemente fáciles de calcular. La media muestral $\bar{y}$ es comúnmente superior a aquéllas debido a que para algunas poblaciones la desviación estándar de su distribución muestral es mínima y siempre es insesgada independientemente de la población.

Algunas propiedades resultan inherentes a la distribución de $\bar{y}$, es decir la distribución muestral generada por muestreo repetido de n observaciones de una población con media μ y varianza σ^2. Se tiene que, independientemente de la distribución de probabilidad de la población, $\bar{y}$ tiene las tres siguientes propiedades.

Características de la distribución muestral de $\bar{y}$

1. El valor esperado de $\bar{y}$ es igual a la media de la población (μ).
2. La desviación estándar de $\bar{y}$ es

$$\sigma_{\bar{y}} = \frac{\sigma}{\sqrt{n}} \sqrt{\frac{N-n}{N-1}}$$

en donde N es el número de observaciones en la población. En la discusión que sigue se supondrá que N es grande en relación con n, el tamaño muestral, y entonces, $\sqrt{(N-n)/(N-1)}$ es aproximadamente igual a 1. Si este es el caso, entonces

$$\sigma_{\bar{y}} = \frac{\sigma}{\sqrt{n}}$$

3. Cuando n es grande, $\bar{y}$ se distribuye aproximadamente como una normal. Esto es consecuencia del teorema central del límite (suponiendo que μ y σ son números finitos).

Entonces $\bar{y}$ es un estimador insesgado de μ, su desviación estándar es proporcional a la desviación estándar de la población, σ, e inversamente proporcional a la raíz cuadrada de n, el tamaño de la muestra. Aunque no se presenta aquí prueba alguna de estos resultados, estos parecen intuitivamente razonables. Por supuesto, entre más variables sean los datos de la población, variabilidad que mide σ, más variable será $\bar{y}$. Por otro lado, cuando n es grande se tiene mayor información para la estimación, y entonces las estimaciones deben estar más cercanas a μ puesto que $\sigma_{\bar{y}}$ decrece al aumentar n.

Además del conocimiento de la media y la desviación estándar de la distribución de probabilidad de $\bar{y}$, el teorema central de límite proporciona información acerca de la forma de esta distribución. Cuando n, el tamaño de muestra, es grande, la distribución de $\bar{y}$ es aproximadamente normal. Esta distribución se exhibe en la figura 8.6.

Para construir un estimador puntual de la media de la población μ, se calcula el valor de la media muestral $\bar{y}$. Entonces, como se explica en la sección 8.4, se calcula la cota para el error de estimación de dos desviaciones estándar. La cota para el error de estimación para la media muestral es

$$\text{cota del error} = \frac{2\sigma}{\sqrt{n}}$$

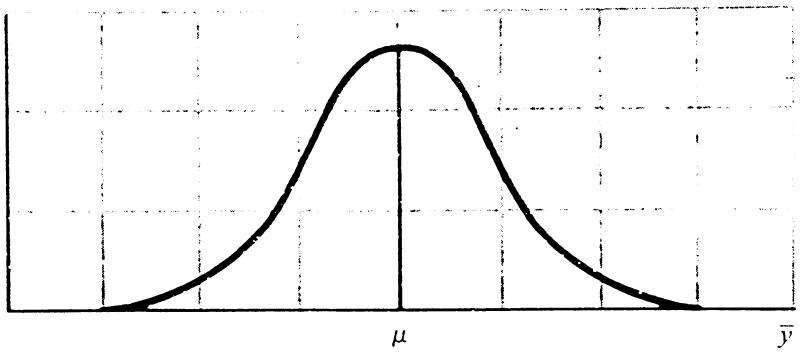


Figura 8.6 *Distribución de $\bar{y}$ para n grande*

La probabilidad de que el error de estimación sea menor que esta cota es aproximadamente .95.

En general se desconoce el valor de σ, la desviación estándar de la población, que aparece en la fórmula de la cota para el error. Entonces es necesario dar una aproximación. Lo más común es substituir σ por s, la desviación estándar de la muestra. Esta aproximación es adecuada cuando el tamaño de muestra es 30 o mayor. Cuando se desconoce σ y el tamaño de muestra es pequeño (menor que 30) debe usarse un procedimiento de estimación con muestras pequeñas. Estos procedimientos se presentan en el capítulo 9.

Estimación puntual de la media de una población

Estimador: $\bar{y}$

Cota del error: $$2\sigma_{\bar{y}} = \frac{2\sigma}{\sqrt{n}}$$

Nota: Si se desconoce σ y n es mayor o igual a 30, se puede usar la desviación estándar muestral s para aproximar σ.

El siguiente ejemplo ilustra el procedimiento de estimación.

Ejemplo 8.1 Suponga que se desea estimar la producción diaria promedio de cierto producto en una planta de productos químicos. Se registra la producción diaria durante $n = 50$ días y se obtiene una media y una desviación estándar de

$$\bar{y} = 871 \text{ tons} \quad \text{y} \quad s = 21 \text{ tons}$$

Estime la producción diaria promedio μ.

Solución La estimación de la producción media es $\bar{y} = 871$ tons. La cota del error de estimación es

$$2\sigma_{\bar{x}} = \frac{2\sigma}{\sqrt{n}} = \frac{2\sigma}{\sqrt{50}}$$

Aunque se desconoce el valor de σ, éste puede ser aproximado por s, el estimador de σ. Se tiene entonces que la cota del error de estimación es aproximadamente

$$\frac{2s}{\sqrt{n}} = \frac{2(21)}{\sqrt{50}} \approx \frac{42}{7.07} \approx 5.94$$

Se puede confiar en que la estimación de 871 tons se encuentra a menos de 5.94 tons del verdadero rendimiento promedio.

El ejemplo 8.1 merece dos comentarios adicionales. Primero, debe evitarse usar el valor equivocado 2σ como cota del error en lugar del valor correcto $2\sigma_{\bar{x}}$, puesto que si se desea discutir acerca de la distribución de $\bar{x}$, debe usarse su desviación estándar $\sigma_{\bar{x}}$ para describir su variabilidad. Se debe tener cuidado para no confundir las medidas descriptivas de una distribución con la de otra.

Segundo, usar s para aproximar σ resulta adecuado cuando n es grande, digamos mayor o igual a 30. Si el tamaño de muestra es pequeño, existen dos posibilidades. En ocasiones se cuenta con una buena estimación de σ a partir de experiencia previa, en tal caso este valor puede usarse en lugar de σ en la fórmula para $\sigma_{\bar{x}}$. Cuando no se cuenta con tal estimación para σ y cuando la muestra se ha seleccionado a partir de una población que tiene una distribución normal (o al menos de forma acampanada) se puede recurrir al procedimiento para muestras pequeñas que se describe en el capítulo 9. El escoger $n = 30$ como punto divisorio entre muestras grandes y pequeñas es arbitrario, y esta elección se discute en el capítulo 9.

Ejercicios

8.1. Debido a la escasez de agua producida por el calor severo del verano en una comunidad, el gobierno de la ciudad selecciona al azar $n = 100$ viviendas para observar el medidor de agua durante un día y estimar el consumo diario promedio por vivienda durante un día caluroso. Se obtiene de esta muestra una media y una desviación estándar de 117.5 galones y 16.8 galones respectivamente. Estime μ, el consumo diario promedio por vivienda en esta comunidad y determine una cota para el error de estimación.

8.2. De acuerdo a una investigación, cuando se intenta evaluar los procedimientos de control interno de una organización mediante un estudio de los libros de contabilidad, para un negocio de tamaño considerable resulta excesivo considerar todas las transacciones efectuadas en un año o aun en un trimestre. Se efectúa una auditoría por medio de una muestra de $n = 50$ transacciones. La media y la desviación estándar que se obtienen son de $2160 y $575. Estime μ, el valor promedio por transacción, y determine una cota del error de estimación.

8.3. La media y la desviación estándar de la utilidad, después de los impuestos, de 100 fabricantes para un período dado de tiempo resultaron de 4.5 y 1.3 centavos por dólar de ingreso. Estime la media de la utilidad después de los impuestos para la población de fabricantes de los cuales se tomó la muestra para ese período de tiempo y determine una cota para el error de estimación.

8.4. Suponga que la media de la población de las utilidades después de los impuestos de todos los fabricantes durante un cierto período es en realidad de 4.8 centavos por dólar de ingreso, con $\sigma = 1.5$ centavos. ¿Cuál es la probabilidad de que la utilidad media, después de los impuestos, calculada de una muestra de $n = 100$ fabricantes exceda 5 centavos?

8.7 Estimación por intervalo de la media de una población

El estimador por intervalo, o **intervalo de confianza,** puede obtenerse fácilmente a partir de los resultados de la sección 8.6. Es posible que $\bar{y}$ pueda resultar mayor o menor que la media de la población aunque no es de esperarse que se desvíe más de $2\sigma_{\bar{y}}$ de μ. Por lo tanto, si se escoge $(\bar{y} - 2\sigma_{\bar{y}})$ como el extremo inferior del intervalo, llamado **límite inferior de confianza** (LIC), y $(\bar{y} + 2\sigma_{\bar{y}})$ como el extremo superior, o **límite superior de confianza** (LSC), el intervalo así construído contendrá muy probablemente a la verdadera media de la población μ. De hecho si n es suficientemente grande como para que la distribución de $\bar{y}$ sea aproximadamente normal, se espera que aproximadamente 95% de los intervalos, que se obtuvieron por muestrear repetidas veces, contendrán a la media de la población μ.

El intervalo de confianza que se describe en el párrafo anterior es llamado **intervalo de confianza de muestras grandes** puesto que se requiere que el tamaño de muestra sea suficientemente grande para que el teorema central del límite garantice la distribución aproximadamente normal de $\bar{y}$. Puesto que comúnmente se desconoce σ **es necesario usar la desviación estándar de la muestra** s **para estimar** σ. **Por lo general este intervalo de confianza resulta apropiado cuando** n **es mayor o igual que 30.**

El coeficiente de confianza de .95 corresponde a $\pm 2\sigma_{\bar{y}}$, o más preciso a $\pm 1.96\sigma_{\bar{y}}$. Ahora, si se recuerda que .90 de las observaciones en una distribución normal se encuentran dentro de $z = 1.645$ desviaciones estándar de la media (tabla 3 del apéndice) se puede ahora construir el intervalo de confianza al 90% con límites

$$\text{LIC} = \bar{y} - 1.645\sigma_{\bar{y}} = \bar{y} - 1.645\frac{\sigma}{\sqrt{n}}$$

$$\text{LSC} = \bar{y} + 1.645\sigma_{\bar{y}} = \bar{y} + 1.645\frac{\sigma}{\sqrt{n}}$$

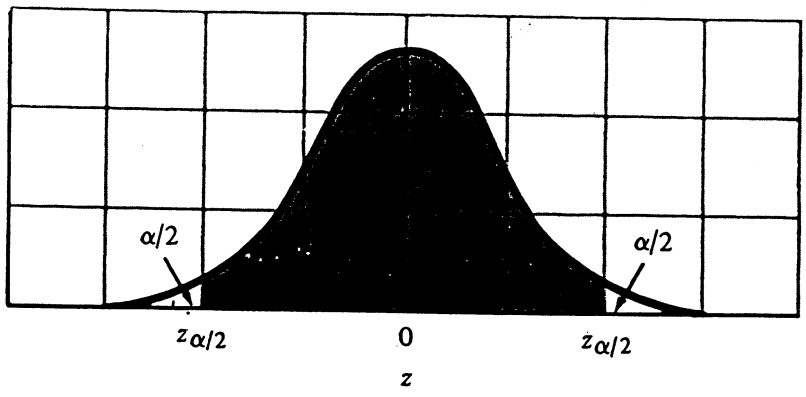


Figura 8.7 *Localización de* $z_{\alpha/2}$

En general, es posible construir intervalos de confianza para cualquier coeficiente de confianza $(1 - \alpha)$ por medio de la fórmula que se presenta en el siguiente cuadro.

Intervalo de confianza del $(1 - \alpha)100\%$ para μ basado en una muestra grande

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

El valor z, $z_{\alpha/2}$, de la curva normal que aparece en la fórmula para el intervalo de confianza, se localiza como se ilustra en la figura 8.7. Por ejemplo si se quiere un coeficiente de confianza de $(1 - \alpha)$ igual a .95, entonces el área requerida bajo las colas de la curva normal es $\alpha = .05$. Se toma la mitad de α (.025) para cada cola de la distribución. Entonces $z_{.025}$ es el valor z de la tabla 3 del apéndice que corresponde a una área de .475 a la derecha de la media. Es decir

$$z_{.025} = 1.96$$

Los límites de confianza correspondientes a los coeficientes de confianza que se usan más frecuentemente se muestran en la tabla 8.1.

Tabla 8.1 *Límites de confianza para μ*

COEFICIENTE DE CONFIANZA $(1 - \alpha)$	α	$z_{\alpha/2}$	LIC	LSC
.90	.10	1.645	$\bar{y} - 1.645\frac{\sigma}{\sqrt{n}}$	$\bar{y} + 1.645\frac{\sigma}{\sqrt{n}}$
.95	.05	1.96	$\bar{y} - 1.96\frac{\sigma}{\sqrt{n}}$	$\bar{y} + 1.96\frac{\sigma}{\sqrt{n}}$
.99	.01	2.58	$\bar{y} - 2.58\frac{\sigma}{\sqrt{n}}$	$\bar{y} + 2.58\frac{\sigma}{\sqrt{n}}$

:mplo 8.2 Encuentre un intervalo de confianza del 90% para la media de la población en el ejemplo 8.1. Recuerde que $\bar{y} = 871$ tons y $s = 21$ tons.

Solución Los límites de confianza para el 90% son:

$$\bar{y} \pm 1.645\frac{\sigma}{\sqrt{n}}$$

Al usar s para estimar σ se tiene

$$871 \pm (1.645)\frac{21}{\sqrt{50}} \approx 871 \pm 4.89$$

Por lo tanto se estima que el rendimiento diario promedio μ cae en el intervalo de 866.11 a 875.89 tons. El coeficiente de confianza .90 implica que en muestreo sucesivo si se determinan los intervalos de confianza para cada muestra, 90% de los intervalos contendrán a μ.

Ejemplo 8.3 Se selecciona una muestra de $n = 100$ empleados de una compañía, se registra el salario anual de cada empleado en la muestra y se calculan la media y la desviación estándar muestral de los salarios obteniéndose

$$\bar{y} = \$7750 \quad \text{y} \quad s = \$900$$

Construya el intervalo de confianza del 95% para el salario promedio de la población μ.

Solución Los límites de confianza al 95% son

$$\bar{y} \pm 1.96 \frac{\sigma}{\sqrt{n}}$$

Usando s como estimador de σ, se tiene

$$\$7750 \pm (1.96) \frac{\$900}{\sqrt{100}} = \$7750 \pm \$176.40$$

Por lo tanto, el salario promedio para los empleados de la compañía está contenido en el intervalo de \$7573.60 a \$7926.40, con coeficiente de confianza de 0.95.

El intervalo de confianza del ejemplo 8.3 es aproximado porque s sustituye como aproximación a σ. Esto es, en lugar de que el coeficiente de confianza sea .95, como se requiere en el ejemplo, el verdadero valor del coeficiente puede ser .92, .94 ó .97. Esto no es para preocuparse desde un punto de vista práctico, puesto que en lo que se refiere a la "confianza" que se tenga en el intervalo hay muy poca diferencia entre estos coeficientes de confianza. La mayoría de los intervalos de confianza usados en estadística producen coeficientes de confianza aproximados puesto que las suposiciones en las que estos se basan no son satisfechas exactamente en la práctica. Habiendo hecho esta aclaración no se volverá a hacer referencia a que los intervalos de confianza son sólo aproximados. En la práctica no hay razón para preocuparse, siempre y cuando el coeficiente de confianza resulte cercano al valor especificado.

Debe notarse en la tabla 8.1 que, para un tamaño de muestra fijo, la longitud del intervalo aumenta al aumentar el coeficiente de confianza, resultado éste que va de acuerdo con la intuición. Si se desea mayor confianza en que el intervalo contrendrá a μ, sería razonable aumentar la amplitud del intervalo. Puesto que se prefieren en general intervalos de confianza no muy amplios con alto coeficiente de confianza es necesario establecer un compromiso al escoger el coeficiente de confianza deseado.

La selección del coeficiente de confianza a usarse en una situación dada debe ser hecha por el experimentador y depende del grado de confianza que el experimentador desea tener en la estimación. La mayoría de los intervalos de confianza se construyen usando uno de los tres coeficientes que aparecen en la tabla 8.1. Tal parece que el más popular de los intervalos de confianza es el del 95%. El uso de intervalos de confianza al 99% es menos frecuente debido a la amplitud del intervalo resultante, aunque es posible siempre reducir la longitud del intervalo al aumentar el tamaño de muestra, n.

El uso tan frecuente del .95 como coeficiente de confianza hace que en ocasiones algunos principiantes se pregunten si debe usarse $z = 1.96$ o $z = 2$ en el intervalo de confianza. La verdad es que no hay mucha diferencia. El valor de $z = 1.96$ es más exacto para un coeficiente de confianza de .95 pero el error que se introduce al usar $z = 2$ es muy pequeño. El uso de $z = 2$ simplifica los cálculos, principalmente cuando las operaciones se realizan manualmente. En general puede acordarse usar dos desviaciones estándar para determinar la cota del error de estimación y usar en cambio $z = 1.96$ para construir intervalos de confianza simplemente para recordar que es el valor obtenido de la tabla de área bajo la curva normal.

Note la diferencia entre los estimadores puntuales y los estimadores por intervalo. Note también que cuando se determina la cota del error de estimación se está construyendo implícitamente un intervalo de confianza, en el caso de que se esté estimando la media de la población. Aunque esta relación se presenta para la mayoría de los parámetros que se estiman en este texto, los dos métodos de estimación no resultan equivalentes. Por ejemplo no resulta obvio que el mejor estimador puntual cae en el centro del mejor estimador por intervalo; en muchos casos no sucede así y más aún no es necesariamente cierto que el mejor estimador por intervalo sea una función del mejor estimador puntual. Aunque estos problemas son de naturaleza teórica, resultan relevantes y deben mencionarse para evitar impresiones equivocadas en el estudiante. Desde un punto de vista práctico, ambos métodos están muy relacionados y la selección entre un estimador puntual y un estimador por el intervalo depende de la preferencia del experimentador.

Ejercicios

8.5. La predicción del volumen de ventas en una región geográfica determinada es una tarea difícil. A menudo se requiere de una estimación de los gastos de los consumidores en una clase especial de productos. En un estudio de mercado realizado por una compañía de productos derivados del petróleo para determinar la cantidad gastada en combustibles para calefacción casera en un año en particular para una ciudad determinada, se utilizó el directorio de la ciudad para seleccionar una muestra aleatoria de $n = 64$ hogares en la ciudad. La media y la desviación estándar muestrales del gasto en combustible para calefacción fueron de \$836 y \$178 respectivamente. Encuentre un intervalo de confianza del 90% para el gasto promedio anual en este tipo de combustible en las viviendas de esa ciudad.

8.6. Para determinar el rendimiento anual de ciertos valores, un grupo de inversionistas tomó una muestra de $n = 50$ de esa clase de valores. La media y desviación estándar resultaron $\bar{y} = 8.71\%$ y $s = 2.1\%$. Estime el verdadero rendimiento anual promedio para esa clase de valores usando un intervalo de confianza del 90%.

8.7. Para llegar a una negociación sindical adecuada se requiere un estimador preciso del salario actual de los empleados sindicalizados. Un agente laboral tomó una muestra de $n = 60$ elevadoristas sindicalizados. En esta muestra se encontró una media y una desviación estándar del salario semanal de los elevadoristas muestreados de \$147.45 y \$11.60 respectivamente. Determine un intervalo de confianza del 95% para el salario semanal promedio de todos los elevadoristas sindicalizados.

8.8. Una auditoría del inventario de una compañía se realizó seleccionando una muestra al azar de $n = 100$ productos en existencia. El precio de venta promedio obtenido en la muestra fue de \$17.50 con una desviación estándar de \$6.75. Construya un intervalo de confianza de 95% para el precio promedio de todos los artículos en existencia.

8.8 Estimación con muestras grandes

La estimación de la media de una población presentada en las seccionas 8.6 y 8.7 ayuda a entender otros problemas de estimación a tratarse en este capítulo. De manera similar a la media muestral el resto de los estimadores presentados en este capítulo siguen una distribución que es aproximadamente normal, muchos de ellos por virtud del teorema central del límite. Todos ellos son estimadores insesgados y las desviaciones estándar de sus distribuciones muestrales son conocidas, por lo que los estimadores puntuales y los intervalos de confianza serán construídos en la misma forma en la que se construyeron para el caso de la media de la población, μ.

En capítulos posteriores se encontrarán también otros estimadores cuyas distribuciones muestrales son aproximadamente normales. Aquí la atención se concentrará en cuatro situaciones que se presentan frecuentemente en la práctica. Los casos que se consideran en este capítulo son inferencias acerca de la media de una población (secciones 8.6 y 8.7), comparación de las medias de dos poblaciones, el parámetro p de una binomial y la comparación del parámetro binomial de dos poblaciones.

Las encuestas que se llevan a cabo tienen frecuentemente el objetivo de estimar la media de la población μ o un parámetro binomial (proporción) p. Por ejemplo, el estimar el número de artículos promedio que desaparecen por robo en un supermercado, es un ejemplo de estimación del promedio de una población μ. La selección aleatoria de una muestra de gente para determinar la proporción de personas a favor de cierto producto casero tiene como objetivo la estimación de un parámetro binomial, p.

Hay también encuestas cuyo objetivo es la comparación de estos parámetros para dos o más poblaciones. Por ejemplo, supóngase que se planea comprar uno de dos automóviles y que se desea escoger aquel que resulte más económico en cuanto a consumo de gasolina. Se sabe además que las cifras de estos rendimientos que se manejan en la publicidad son bastante alejados de la realidad y que el kilometraje por litro varía considerablemente de un automóvil a otro. La decisión puede basarse en una comparación de las medias de muestras aleatorias de los dos tipos de automóvil, n_1 del primer tipo y n_2 del segundo tipo. La decisión se basará en la diferencia de los rendimientos promedio de los dos tipos de automóvil. De manera similar el objetivo puede ser estimar la diferencia de las proporciones de productos defectuosos para dos grandes lotes de productos industriales, la estimación basada en muestras aleatorias seleccionadas de cada lote.

Las situaciones descritas anteriormente identifican los cuatro problemas de estimación que se presentan en este capítulo en que se presentan estimadores para las siguientes situaciones:

1. La media de una población μ.
2. La diferencia entre las medias de dos poblaciones.
3. La proporción de una población.
4. La diferencia entre las proporciones de dos poblaciones.

Como se estableció anteriormente, las distribuciones muestrales de estos estimadores, cuando el tamaño de muestra es grande, poseen propiedades similares, por lo que se puede dar una forma general para la estimación puntual y por intervalos para los cuatro problemas de estimación. Aunque las formas del estimador puntual defieren, la cota del error de cada uno se presenta en el siguiente cuadro:

Cota para el error de estimación para un estimador puntual con una muestra grande

Cota del error = 2 desviaciones estándar de la distribución muestral del estimador puntual.

Similarmente los intervalos de confianza con muestras grandes para los cuatro parámetros se presentan en el cuadro siguiente:

Intervalo de confianza con muestras grandes del $(1 - \alpha)100\%$

Estimador puntual $\pm z_{\alpha/2} \times$ (desviación estándar del estimador puntual) donde $z_{\alpha/2}$ se obtiene de la tabla 8.1.

Las fórmulas específicas para la estimación de la diferencia de dos medias, para la estimación de una proporción binomial y para la estimación de la diferencia de dos proporciones se presentan en las secciones 8.9, 8.10 y 8.11.

8.9 Estimación de la diferencia de dos medias

Un problema igualmente importante al de la estimación de una media poblacional es el de la estimación de la diferencia entre dos medias poblacionales. Por ejemplo, podría requerirse estimar el tiempo medio requerido para el ensamblado de un producto industrial usando dos métodos alternativos. Los armadores se dividirían en forma aleatoria en dos grupos, uno de los cuales procedería al ensamblado utilizando el método 1 y el otro el método 2. Se podría entonces, hacer inferencias en relación a la diferencia entre los tiempos medios requeridos por los dos métodos. En otro ejemplo similar, suponga que en una planta química, se quiere comparar el rendimiento medio utilizando materia prima de dos proveedores alternativos, A o B. Si se toman muestras diarias de los rendimientos utilizando materia prima de A por un lado y de B por el otro, estos pueden ser utilizados para inferir sobre la diferencia entre los rendimientos medios.

En ambos ejemplos, hay dos poblaciones, la primera con media y varianza μ_1 y σ_1^2, respectivamente, y la segunda con media y varianza μ_2, σ_2^2. Una muestra aleatoria de tamaño n_1 se extrae de la población 1 y otra de tamaño n_2 de la población 2, en donde se supone que las muestras han sido extraídas de manera independiente la una de la otra. Sean $\bar{y}_1$, s_1^2, $\bar{y}_2$ y s_2^2 los estimadores de los parámetros, calculados con la información muestral.

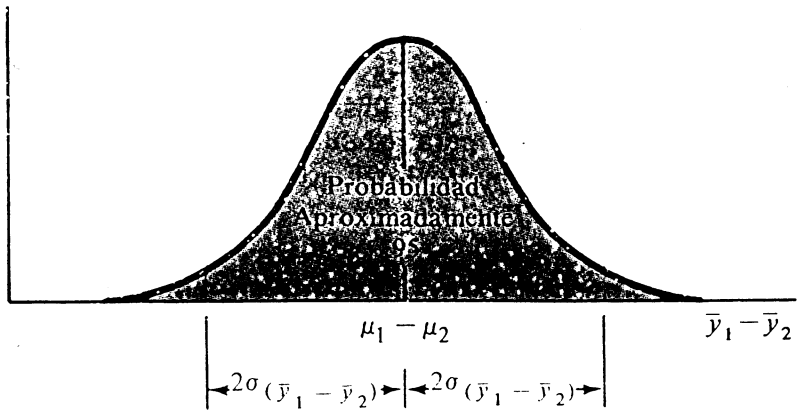


Figura 8.8 *La distribución de* $(\bar{y}_1 - \bar{y}_2)$ *para muestras grandes*

El estimador puntual de la diferencia $(\mu_1 - \mu_2)$ de las medias poblaciones, es $(\bar{y}_1 - \bar{y}_2)$, la diferencia entre las medias muestrales. Si se repitiese el muestreo y en cada ocasión se calculase $(\bar{y}_1 - \bar{y}_2)$, resultaría una distribución del estimador. ¿Cuáles son las propiedades de esta distribución muestral? Se puede demostrar que la distribución de la suma o la diferencia entre dos medias muestrales normales e independientes, es normal, y en el caso de la diferencia, con media $(\mu_1 - \mu_2)$.

La desviación estándar para esta distribución muestral es

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

en donde σ_1^2 y σ_2^2 son las varianzas poblacionales y n_1, n_2 los tamaños de muestra respectivos. La distribución muestral se ilustra en la figura 8.8.

Un resumen de lo anterior se presenta en el siguiente cuadro.

Estimación puntual de $(\mu_1 - \mu_2)$

Estimador: $(\bar{y}_1 - \bar{y}_2)$.

Cota para el error: $2\sigma_{(\bar{y}_1 - \bar{y}_2)} = 2\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}$.

Nota: Si σ_1^2, σ_2^2 son desconocidas, pero n_1, n_2 son mayores o iguales que 30, se pueden utilizar s_1^2, s_2^2 en lugar de σ_1^2, σ_2^2.

Se ilustra el uso de este procedimiento con un ejemplo.

Ejemplo 8.4 Una comparación del desgaste de dos tipos distintos de neumáticos para automóvil, se hizo rodando $n_1 = n_2 = 100$ neumáticos de cada tipo. El número de kilómetros-vida de cada neumático se anotó, en donde km-vida fue definido como el kilometraje andado antes de que el neumático quedase en un estado determinado de desgaste. Los resultados de la prueba fueron los siguientes:

NEUMÁTICO 1	NEUMÁTICO 2
$\bar{y}_1 = 26{,}400$ km	$\bar{y}_2 = 25{,}100$ km
$s_1^2 = 1{,}440{,}000$	$s_2^2 = 1{,}960{,}000$

Estime $(\mu_1 - \mu_2)$, la diferencia en km-vida medios, y obtenga la cota para el error de estimación.

Solución El estimador puntual de $(\mu_1 - \mu_2)$ es

$$(\bar{y}_1 - \bar{y}_2) = 26{,}400 - 25{,}100 = 1{,}300 \text{ km}$$

Se tiene entonces,

$$\sigma_{(\bar{y}_1 - \bar{y}_2)} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \approx \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

$$= \sqrt{\frac{1{,}440{,}000}{100} + \frac{1{,}960{,}000}{100}} = \sqrt{34{,}000} \approx 184 \text{ km}$$

Se esperará que el error de estimación sea menor que $2(184) = 368$ km-vida. Si el estimador de la diferencia en km-vida medios es 1,300 y el error de estimación es menor que 368 (con una alta probabilidad), parece razonable concluir que hay una diferencia sustancial entre las medias de km-vida de los dos tipos de neumáticos. De hecho, el neumático 1 es el que se muestra superior al 2 en cuanto a desgaste por rodamiento.

Siguiendo el prodedimiento de la sección 8.7, se puede construir un intervalo de confianza con coeficiente de confianza $(1 - \alpha)$ para $(\mu_1 - \mu_2)$.

Un intervalo de confianza del $(1 - \alpha)$ 100% para $(\mu_1 - \mu_2)$.

$$(\bar{y}_1 - \bar{y}_2) \pm z_{\alpha/2} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

Nota: Si n_1, n_2 son mayores o iguales que 30, se pueden utilizar s_1^2 s_2^2 en lugar de σ_1^2, σ_2^2.

Ejemplo 8.5 Establezca un intervalo de confianza para la diferencia de los km-vida medios al problema descrito en el ejemplo 8.4. Utilice un coeficiente de confianza del .99.

Solución El intervalo de confianza es

$$(\bar{y}_1 - \bar{y}_2) \pm 2.58 \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

Substituyendo los resultados del ejercicio 8.4 el intervalo resultante es

$$1{,}300 \pm 2.58(184)$$

En otras palabras, LIC $\approx$ 825, LSC $\approx$ 1,775 y la diferencia de km-vida medios se estima que se encuentra entre estos dos límites con una confiabilidad del .99. Note que este intervalo es más grande que el de $\pm 2\sigma_{(\bar{y}_1 - \bar{y}_2)}$ utilizado en el ejemplo 8.4, ya que en éste se ha escogido un coeficiente de confianza mayor.

Ejercicios

8.9. Un nuevo reglamento sobre la emisión máxima de ruido permisible en camiones y autobuses de más de cinco tons en tráfico interestatal, limita el número de decibeles a 90, medidos a 15 metros, en zonas con velocidad máxima superior a 60 km/h. En zonas de menor velocidad, el número de decibeles tolerado es menor. Con el propósito de estudiar si durante la instrumentación del nuevo reglamento se tendrá que enfatizar más el control entre el gremio de transportistas cargueros o el de compañías de autobuses, se condujo un experimento consistente en la observación, a 15 metros, del ruido emitido por 50 camiones y 50 autobuses en zonas de velocidad máxima superior a 60 km/h. La media muestral del número de decibeles y la desviación estándar muestral para los camiones fueron 91.5 y 5.2 respectivamente, y para los autobuses 88.6 y 3.8. Encuentre un intervalo de confianza del 95% para la diferencia entre las emisiones medias de ruido de camiones y de autobuses en las circunstancias descritas.

8.10. Casi todas las escuelas de posgrado requieren de la aprobación de un exámen de admisión que mide tanto la habilidad verbal como la quantitativa de los candidatos. Suponga que uno de estos exámenes, con escala 0-800, se utiliza para comparar a los estudiantes ya admitidos en dos escuelas distintas. Se seleccionan dos muestras aleatorias consistentes en 75 calificaciones correspondientes a 75 estudiantes de cada una de las escuelas. Los promedios muestrales y desviaciones estándar calculados son:

escuela 1: $\bar{y}_1 = 525$ $s_1 = 52$

escuela 2: $\bar{y}_2 = 562$ $s_2 = 70$

Estime la diferencia entre la calificación media de los estudiantes de cada una de las escuelas y dé la cota para el error de estimación.

8.11. Las legislaciones en torno a la protección al consumidor, entre otras cosas, han ocasionado que las empresas de manufactura se preocupen más por la aceptación de sus productos en el mercado. Una empresa, con dos productos en su línea, quiere estimar la diferencia entre el número de quejas ocurridas cada mes durante cuatro años (48 meses), que para cada productos arrojan los siguientes resultados:

PRODUCTO 1	PRODUCTO 2
$\bar{y}_1 = 17.2$	$\bar{y}_2 = 25.1$
$s_1 = 4.6$	$s_2 = 5.3$

Encuentre un intervalo de confianza del 90% para la diferencia entre el número medio de quejas sobre los dos productos (suponga que las quejas sobre cada producto se pueden considerar como muestras aleatorias independientes).

8.10 Estimación del parámetro de una población binomial

Como se ha hecho notar, el objetivo de muchas encuestas es la estimación de la proporción de personas u objetos que poseen un atributo particular. Una muestra de esta naturaleza es un ejemplo práctico del experimento binomial discutido en el capítulo 6. La estimación de la proporción de las ventas que pueden esperarse al establecer contacto con un gran número de clientes es un

ejemplo de un problema práctico que requiere la estimación de un parámetro binomial p.

El mejor estimador de un parámetro binomial p es, de nuevo, aquel que se hubiera escogido intuitivamente. Esto es, el estimador de p, que se denotará por el símbolo $\hat{p}$, es el número total de éxitos y, dividido entre el número total de ensayos, n, es decir:

$$\hat{p} = \frac{y}{n}$$

en donde y es el número de éxitos en n ensayos. (*Nota:* el tilde ($\hat{}$) sobre el parámetro es el símbolo que se usa para denotar el estimador del parámetro.) Cuando se dice que $\hat{p}$ es el "mejor" estimador significa que $\hat{p}$ es insesgado y que su varianza es menor que la de cualquier otro posible estimador insesgado.

Como se señaló en la sección 8.8, debido al teorema central del límite, el estimador $\hat{p}$ posee una distribución muestral que se aproxima, para n grande, de manera adecuada a la distribución normal. Resulta que $\hat{p}$ es insesgado para la proporción de la población p y su desviación estándar está dada por

$$\sigma_{\hat{p}} = \sqrt{\frac{pq}{n}}$$

donde p es la proporción desconocida de la población, $q = (1 - p)$, y n es el tamaño de muestra. Entonces de acuerdo a la sección 8.8 se dá en el siguiente cuadro el procedimiento para la estimación puntual de p.

Estimación puntual para p

Estimador: $\hat{p} = \frac{y}{n}$.

Cota del error: $2\sigma_{\hat{p}} = 2\sqrt{\frac{pq}{n}}$

A continuación se presenta el intervalo de confianza con base en muestras grandes con coeficiente de confianza $(1 - \alpha)$.

Intervalo de confianza para p del $(1 - \alpha)100\%$

$$\hat{p} \pm z_{\alpha/2}\sqrt{\frac{pq}{n}}$$

Se considera que el tamaño de muestra es grande cuando se puede suponer que la distribución de $\hat{p}$ se aproxima adecuadamente a la distribución normal. Estas condiciones fueron discutidas en la sección 7.5.

La única dificultad que se encuentra en el procedimiento es el cálculo de $\sigma_{\hat{p}}$ que requiere usar p y $q = (1 - p)$ que son desconocidos. Como se verá se habrá de substituir $\hat{p}$ en lugar del parámetro p en la fórmula para $\sigma_{\hat{p}}$. Cuando n es grande, el error que se comete por esta substitución es bastante pequeño. En realidad la desviación estándar $\sqrt{pq/n}$ cambia sólo ligeramente cuando p varía. Este hecho puede ser observado en la tabla 8.2 en donde se exhibe

Tabla 8.2 *Algunos valores de* $\sqrt{pq}$

p	$\sqrt{pq}$
.5	.50
.4	.49
.3	.46
.2	.40
.1	.30

$\sqrt{pq}$ para diversos valores de p. Nótese que $\sqrt{pq}$ cambia muy poco cuando p cambia, especialmente cuando p es cercano a .5.

Ejemplo 8.6 Una muestra aleatoria de $n = 100$ electores en una comunidad produjo $y = 59$ electores en favor del candidato A. Estime la fracción de electores a favor del candidato A y determine una cota para el error de estimación.

Solución El estimador puntual de p es

$$\hat{p} = \frac{y}{n} = \frac{59}{100} = .59$$

y la cota para el error de estimación es:

$$2\sigma_{\hat{p}} = 2\sqrt{\frac{\hat{p}\hat{q}}{n}} = 2\sqrt{\frac{(.59)(.41)}{100}} = .098$$

El intervalo de confianza del 95% para p es

$$\hat{p} \pm 1.96\sqrt{\frac{\hat{p}\hat{q}}{n}} = .59 \pm 1.96(.049)$$

Entonces se estima que el intervalo de .494 a .686 contiene a p con un coeficiente de confianza del .95.

Ejercicios

8.12. Una liga nacional de sindicatos obreros necesita estimar la proporción de obreros no sindicalizados que estarían a favor de una huelga nacional. Sesenta de 200 obreros entrevistados en centros laborales no sindicalizados se encontraron a favor de la huelga. Estime la proporción de obreros no sindicalizados que están a favor de la huelga y determine una cota para el error de estimación.

8.13. Generalmente los estudios de factibilidad de proyectos requieren de una medida de la demanda para determinar la redituabilidad potencial de un bien o servicio. En un estudio para determinar la factibilidad de aumentar la programación de televisión con apoyo gubernamental, un investigador encontró que 76 de 180 viviendas con televisor seleccionadas totalmente al azar ven programas con apoyo gubernamental al menos dos horas a la semana. Encuentre usted un intervalo de confianza del 90% para la proporción p de viviendas con televisor que ven al menos 2 horas a la semana de programas patrocinados por el gobierno.

8.14. En los Estados Unidos, en 1975, apareció en diversos periódicos el siguiente artículo:

Casi 30% de los norteamericanos no conocen la importancia histórica de 1776

Casi a punto de celebrarse el aniversario 200, casi 3 de 10 norteamericanos son incapaces de decir qué evento histórico ocurrió en 1776. En una muestra aleatoria de 550 personas se encontró que solamente 396 de los entrevistados resultaron capaces de identificar a 1776 como el año en el que Estados Unidos declaró su independencia de la Gran Bretaña.

Con base en esta información estime la proporción de norteamericanos que pueden identificar la importancia de 1776 y determine una cota para el error de estimación.

8.15. En una encuesta reciente de la agencia Gallup se observó que en una muestra de 1,200 residentes de la ciudad de Sao Paulo, Brazil, el 82% de los entrevistados consideran la contaminación de la atmósfera como un "serio problema." Encuentre el intervalo de confianza del 95% para la proporción de la población de Sao Paulo que consideran la contaminación del aire como un problema serio.

8.16. Una encuesta mostró que el 20% de una muestra aleatoria de $n = 1500$ consumidores en una ciudad dada piensan adquirir automóviles nuevos en el próximo año. Encuentre un intervalo de confianza del 90% para la proporción p de habitantes de la ciudad que planean adquirir un auto nuevo el año próximo.

8.11 Estimación de la diferencia entre 2 proporciones

El cuarto, y último problema de estimación que se considera en este capítulo es la estimación de la diferencia entre los parámetros de dos poblaciones binomiales. Supóngase que las poblaciones 1 y 2 poseen parámetros p_1 y p_2 respectivamente y que se toman muestras aleatorias independientes de n_1 y n_2 ensayos respectivamente y se calculan estimaciones $\hat{p}_1$ y $\hat{p}_2$.

Se notó ya en la sección 8.10 que la proporción muestral,

$$\hat{p} = \frac{y}{n} ,$$

de una muestra aleatoria de n elementos de una población binomial, posee una distribución muestral que es aproximadamente normal cuando n es grande. Como consecuencia del teorema central de límite entonces, al igual que en el caso de la diferencia de dos medias muestrales, la diferencia en las proporciones muestrales

$$(\hat{p}_1 - \hat{p}_2)$$

sigue una distribución muestral que es aproximadamente normal cuando ambos n_1 y n_2 son grandes. Además se establece aquí, sin demostración, que el estimador $(\hat{p}_1 - \hat{p}_2)$ es un estimador insesgado de la diferencia $(p_1 - p_2)$ de las proporciones de la población y posee una desviación estándar dada por

$$\sigma_{(\hat{p}_1 - \hat{p}_2)} = \sqrt{\frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}}$$

donde p_1 y p_2 son las proporciones en las dos poblaciones, $q_1 = (1 - p_1)$, $q_2 = (1 - p_2)$, y n_1 y n_2 los tamaños de muestra seleccionada en cada una de las dos poblaciones.

Usando el procedimiento de la sección 8.8 se presenta a continuación el estimador para $(p_1 - p_2)$.

Estimación puntual de $(p_1 - p_2)$

Estimador: $(\hat{p}_1 - \hat{p}_2)$

Cota del error: $2\sigma_{(\hat{p}_1 - \hat{p}_2)} = 2\sqrt{\dfrac{p_1 q_1}{n_1} + \dfrac{p_2 q_2}{n_2}}$

Nota: los estimadores $\hat{p}_1$ y $\hat{p}_2$ son usados en lugar de p_1 y p_2 para obtener la cota del error de estimación.

El intervalo de confianza del $(1 - \alpha)$ para usarse cuando n_1 y n_2 son ambos grandes se exhibe en el siguiente cuadro.

Intervalo de confianza del $(1 - \alpha)100\%$ para $(p_1 - p_2)$

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

Ejemplo 8.7 Un fabricante de insecticidas en presentación aerosol desea comparar dos productos nuevos, 1 y 2. Se emplean en el experimento dos cuartos del mismo tamaño, cada uno con 1000 moscas. En uno se rocía el insecticida 1 y en el otro se rocía el insecticida 2 en igual cantidad. Se obtienen totales de 825 y 760 moscas muertas por acción de los insecticidas 1 y 2 respectivamente. Estime la diferencia de la proporción de éxitos para los dos insecticidas cuando se usan en condiciones similares a las probadas.

Solución El estimador puntual de $(p_1 - p_2)$ es

$$(\hat{p}_1 - \hat{p}_2) = .825 - .760 = .065$$

La cota para el error de estimación es de

$$2\sqrt{\frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}} \approx 2\sqrt{\frac{(.825)(.175)}{1000} + \frac{(.76)(.24)}{1000}} = .036$$

El intervalo de confianza del 95% es

$$(\hat{p}_1 - \hat{p}_2) \pm 1.96\sqrt{\frac{\hat{p}_1\hat{q}_1}{n_1} + \frac{\hat{p}_2\hat{q}_2}{n_2}}$$

por lo que el intervalo resultante es

$$.065 \pm .035$$

Por lo tanto se estima que la diferencia $(p_1 - p_2)$ de proporciones de éxito está entre .030 y .100. Esto es, se estima que p_1 excede a p_2 por al

menos .030 y a lo más por .100. Se puede tener buena confianza en esta estimación, puesto que, como se ha visto, si se repitiera el mismo procedimiento de muestreo una y otra vez y para cada vez se determinara el intervalo de confianza, aproximadamente 95% de los intervalos contendrían la cantidad $(p_1 - p_2)$.

Ejercicios

8.17. En una encuesta realizada entre los accionistas de una compañía 300 de 500 hombres estuvieron a favor de lanzar una nueva línea de productos, mientras 64 de 100 mujeres apoyaron el proyecto. Estime la diferencia de las proporciones de personas que favorecen la decisión y determine una cota para el error de estimación.

8.18. La toma de decisiones participativa ha sido una estrategia administrativa que se ha adoptado como un medio para mejorar la eficiencia y la participación de los individuos en las organizaciones. Se entrevistó a dos grupos de empleados, los cuales difieren substancialmente en el nivel de participación permitida por el patrón, y se les preguntó si estaban o no satisfechos con su empleo actual. De 110 empleados de un grupo en el cual se ha fomentado la participación del empleado, 77 afirmaron que estaban satisfechos con sus empleos. En tanto 52 de 125 empleados, de un grupo en el que no se permite la participación del empleado, afirmaron que estaban satisfechos con su empleo.

a. Estime la diferencia en la proporción de empleados satisfechos con sus trabajos, y determine una cota para el error de estimación.

b. Encuentre un intervalo de confianza del 90% para la diferencia de la proporción de empleados satisfechos con su trabajo.

8.19. Tomeski y Lazarus (1974) reportan acerca del uso de los sistemas de información computarizados para el manejo de registros de personal en los gobiernos estatales y locales en los Estudios Unidos.* Suponga que en un estudio similar se tiene que 33 de 40 gobiernos de ciudades seleccionadas usan sistemas de información computarizados para el manejo de sus registros de personal, mientras 25 de 32 gobiernos estatales usan esos sistemas. Estime la diferencia en las proporciones de gobiernos de las ciudades y gobiernos estatales en los Estados Unidos que usan sistemas de información computarizados para los registros de personal. Use un intervalo de confianza del 95%.

*Tomeski y Lazarus, "Computerized Information Systems in Personnel," *Academy of Management Journal*, marzo de 1974.

8.12 Selección del tamaño de muestra

El diseño de un experimento es esencialmente un plan para conseguir cierta cantidad de información. Esta información, como cualquier otro bien o servicio puede ser adquirida a precios que varían dependiendo de la forma en que sean obtenidos los datos. Algunas de las observaciones contienen una gran cantidad de información acerca de los parámetros de interés, otras pueden proporcionar escasa o nula información. Puesto que el producto directo de la investigación es la información, ésta debe ser adquirida al menor costo posible.

El procedimiento de muestreo o el diseño experimental, como se le llama comúnmente, influye en la cantidad de información por observación. Este, junto con el tamaño de la muestra n, determina la cantidad total de información relevante en la muestra. Salvo en contadas ocasiones, en este libro se tratará con la más simple de las situaciones de muestreo, muestreo aleatorio simple, o totalmente al azar, de una población relativamente grande, así que la atención será enfocada solamente a la selección del tamaño de muestra.

Poco es lo que puede avanzarse en la planeación de un experimento antes de enfrentarse al problema de determinar el tamaño de la muestra. Más aún quizá la pregunta que los estadísticos oyen más a menudo es, ¿cuántas observaciones deben ser incluídas en la muestra? Por desgracia el estadístico no puede responder a esta pregunta sin saber antes cuanta información el experimentador desea adquirir. Lo que es seguro es que la cantidad de información total en la muestra afectará la bondad del método de inferencia a utilizar y debe ser especificada por el experimentador. En lo que se refiere específicamente al problema de estimación, la precisión deseada puede ser determinada al fijar una cota para el error de estimación.

Por ejemplo, suponga que se desea estimar el rendimiento diario promedio μ de un producto químico (ver ejemplo 8.1) y se desea que el error de estimación sea menor que 4 tons, con una probabilidad de .95. Puesto que aproximadamente 95% de las medidas muestrales caerán dentro de $2\sigma_{\bar{y}}$ de distancia de μ al repetir el procedimiento de muestreo una y otra vez, lo que se está pidiendo es $2\sigma_{\bar{y}} = 4$ (ver figura 8.9). Entonces

$$2\sigma_{\bar{y}} = 4 \qquad \text{o} \qquad \frac{2\sigma}{\sqrt{n}} = 4$$

y resolviendo para n se tiene:

$$n = \frac{\sigma^2}{4}$$

Este es el tamaño de muestra mínimo tal que el error de estimación será menor que $2\sigma_{\bar{y}}$ con probabilidad de .95.

Como se habrá notado no es posible encontrar un valor número para n a menos que la desviación estándar de la población σ sea conocida. Y esto es precisamente lo que era de esperarse puesto que la variabilidad de la media muestral $\bar{y}$ depende de la variabilidad de la población de la cual la muestra fue tomada. A falta del valor exacto de σ, se deberá usar la mejor aproximación disponible. Este puede ser el estimador s obtenido de una encuesta previa o puede usarse el conocimiento de la amplitud en la cual las observaciones caen. Puesto que la amplitud es aproximadamente de 4σ (la regla empírica), un cuarto de la amplitud proporciona un valor aproximado para σ. Para este

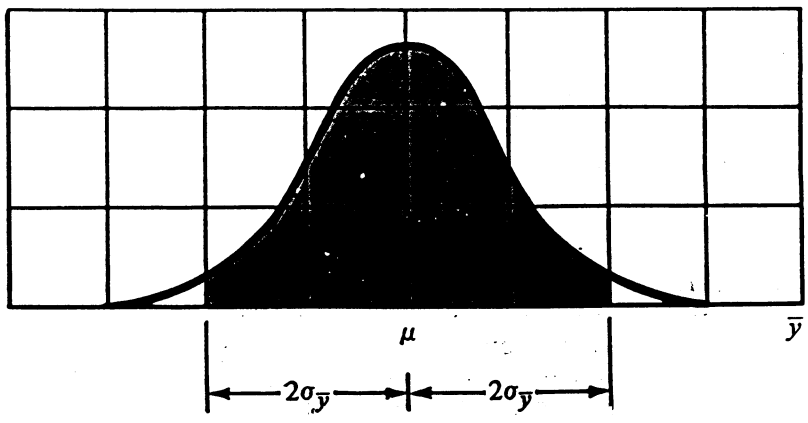


Figura 8.9 *Distribución aproximada de $\bar{y}$ para muestras grandes*

ejemplo se usan los resultados del ejemplo 8.1 que proporcionan una estimación bastante precisa de σ, $s = 21$. Entonces:

$$n = \frac{\sigma^2}{4} \approx \frac{(21)^2}{4} = 110.25 \approx 111$$

Si en este caso se usa un tamaño de muestra $n = 111$ se estará razonablemente seguro (con probabilidad aproximada de .95) de que el estimador caerá dentro de $2\sigma_{\bar{Y}} = 4$ tons de distancia de μ, el verdadero rendimiento promedio diario.

De hecho se espera que el error de estimación sea menor que 4 tons. De acuerdo a la regla empírica se tiene una probabilidad aproximada de .68 de que el error de estimación sea menor que $\sigma_{\bar{Y}} = 2$ tons. Debe notarse que las probabilidades de .68 y .95 son aproximadas puesto que σ fue substituído por s. Aunque este método para seleccionar el tamaño de muestra es sólo aproximado para una precisión deseada en la estimación, es el mejor disponible, y resulta por supuesto mejor que seleccionar el tamaño de muestra basándose solamente en la intuición.

El método para seleccionar el tamaño de muestra para todos los procedimientos de estimación con muestras grandes discutidos en las secciones anteriores es muy similar a como se describe anteriormente. El experimentador debe especificar una cota para el error de estimación y un coeficiente de confianza asociado $(1 - \alpha)$.

Procedimiento para seleccionar el tamaño de muestra

Sea θ (la letra griega *theta*) el parámetro que se desea estimar y sea $\sigma_{\hat{\theta}}$ *la desviación estándar del estimador puntual para* θ. El procedimiento es:

1. Escoja B, la cota para el error de estimación y un coeficiente de confianza $(1 - \alpha)$.
2. Resuelva la ecuación siguiente para obtener el tamaño de muestra n:

$$z_{\alpha/2}\ \sigma_{\hat{\theta}} = B$$

Nota: Para la mayoría de los estimadores (todos los que se usan en este texto) $\sigma_{\hat{\theta}}$ *es una función del tamaño de la muestra* n.

Las ideas anteriores se ilustran ahora con dos ejemplos.

Ejemplo 8.8 Uno de los mayores problemas que enfrentan las organizaciones que dependen de la información de los consumidores para diseñar sus políticas es la dificultad de seleccionar consumidores que respondan al cuestionario de la encuesta. Chervany y Heinlen reportan su experiencia con encuestas piloto para determinar tasas de respuesta para diferentes versiones de un cuestionario.*

Suponga que un ejecutivo de la organización desea estimar la tasa de respuesta para un cuestionario recientemente diseñado y quiere basar su estimación en la tasa de respuesta obtenida en una encuesta piloto. ¿Cuántos

*N. L. Chervany and J. S. Heinlen, "The Structure of a Student Project Course," *Decision Sciences*, enero de 1975.

consumidores deben seleccionarse en la encuesta piloto si el ejecutivo desea estimar p, la tasa de respuesta, con una probabilidad de .90 de que el error de estimación sea menor que .06? Suponga que se espera que p esté cercano a .6.

Solución Puesto que el coeficiente de confianza es $(1 - \alpha) = .90$, α debe ser igual a .10 y $\alpha/2 = .05$. El valor de $z_{\alpha/2}$ que corresponde a una área de .05 en la cola superior de la distribución normal estándar es $z_{\alpha/2} = 1.645$. Se requiere entonces

$$1.645\sigma_{\hat{p}} = .06 \qquad \text{o} \qquad 1.645\sqrt{\frac{pq}{n}} = .06$$

Puesto que la variabilidad de $\hat{p}$ depende de p, que es desconocida, se debe usar el valor de .6 predicho por el ejecutivo como una aproximación. Entonces

$$1.645\sqrt{\frac{(.6)(.4)}{n}} = .06$$

resolviendo para n se tiene

$$\sqrt{n} = \frac{1.645}{.06}\sqrt{(.6)(.4)} \approx 13.43$$

$$n \approx 180.4 \approx 181$$

Es decir la encuesta piloto debe incluir 181 consumidores.

Ejemplo 8.9 Un experimentador desea comparar la efectividad de dos métodos de entrenamiento industrial para obreros que han de llevar a cabo una operación especial en una planta armadora. Los empleados seleccionados se dividen en dos grupos. El primer grupo recibe el método 1, el segundo el método 2. Cada uno realizará la operación de armado y se registra el tiempo que emplea en realizar la operación. Se espera que las observaciones en ambos grupos tengan un rango de 8 minutos. Si se desea que el estimador de la diferencia en tiempo medio de armado sea correcta hasta por un minuto, con probabilidad aproximada de .95, ¿cuántos trabajadores han de incluírse en cada grupo?

Solución Igualando $2\sigma_{(\bar{y}_1 - \bar{y}_2)}$ a un minuto, se tiene

$$2\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} = 1$$

Puesto que se desea n_1 igual a n_2, sea $n_1 = n_2 = n$ para obtener la ecuación:

$$2\sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{n}} = 1$$

Se espera que la variabilidad de cada método de armado sea la misma, y entonces $\sigma_1 = \sigma_2 = \sigma$. Puesto que la amplitud es igual a 8 minutos que es aproximadamente igual a 4σ, entonces se tiene

$$4\sigma \approx 8$$

$$\sigma \approx 2$$

al substituir este valor para σ_1 y σ_2 se obtiene

$$2\sqrt{\frac{(2)^2}{n} + \frac{(2)^2}{n}} = 1$$

$$\sqrt{n} = 2\sqrt{8}$$

o $n = 32$. Cada grupo debe contener $n = 32$ empleados.

Ejercicios

8.20. Los productos defectuosos resultan costosos para el fabricante en términos de costo por reemplazo y en términos de deterioro de la imagen del producto ante el público consumidor. Un fabricante de tocacintas portátiles confía en que a lo más 10% de los productos fabricados por su empresa resultan defectuosos. Si se desea estimar la proporción actual de tocacintas defectuosos dentro de un .03, ¿cuántos tocacintas deben ser seleccionados en la muestra?

8.21. El mantenimiento de cuentas de crédito puede resultar demasiado costoso si el promedio de compra por cuenta baja de cierto nivel. El gerente de un almacén desea estimar el promedio de cantidad comprada por mes por sus clientes que usan cuenta de crédito, con un error de no más de $2.50 con una probabilidad aproximada de .95. ¿Cuántas cuentas deben ser seleccionadas del archivo de la compañía si se sabe que la desviación estándar de los balances mensuales de las cuentas de crédito es de $7.50?

8.22. Para poder negociar competitivamente los derechos de explotación de una zona forestal, una compañía maderera necesita conocer el diámetro promedio de los árboles en la zona con una precisión de menos de 2 pulgadas, con una probabilidad de .90. ¿De qué tamaño debe ser la muestra de árboles seleccionados para estimar el diámetro promedio, si se sabe que los diámetros de madera utilizable en la zona varían de 10 a 38 pulgadas?

8.23. ¿Es el costo de operación de los autos subcompactos notablemente menor que el de los autos compactos? De acuerdo a un artículo en el *U.S. News and World Report,** la respuesta es no necesariamente. Para investigar este tema una agencia privada desea estimar la diferencia en el costo promedio de operación por milla entre sub-compactos y compactos con una precisión de menos de 1 centavo por milla, con probabilidad aproximada de .95. Si se sabe por experiencia que la desviación estándar de los costos de operación es de 3 centavos, ¿cuántos autos han de considerarse en cada grupo si se desea usar el mismo número de autos en cada grupo?

*"How much Does Car Ownership Really Cost?" *U.S. News and World Report*, noviembre de 1975.

8.13 Prueba estadística de una hipótesis

El razonamiento básico empleado en la prueba estadística de una hipótesis ha sido bosquejado en la sección 6.8 al discutir una prueba de la efectividad de una vacuna contra el resfriado. En esta sección se hará una condensación de los puntos básicos involucrados en el problema y se sugiere repasar la sección 6.8 para recordar la presentación intuitiva del tema.

Partes de una prueba estadística

El objetivo de una prueba estadística es probar una hipótesis acerca de uno o más parámetros de una población. En una prueba estadística se encuentran involucrados los siguientes cuatro elementos:

1. Hipótesis nula.
2. Hipótesis alternativa.
3. Estadística de prueba.
4. Región de rechazo.

Al especificar estos cuatro elementos queda definida una prueba en particular; y al cambiar una o más de estas partes se produce una prueba diferente.

La relación entre las hipótesis nula y alternativa se discutió en la sección 6.9. La **hipótesis alternativa** o hipótesis de investigación as aquella que el investigador desea apoyar. La **hipótesis nula** es la contradicción de la hipótesis alternativa; esto es, si la hipótesis nula es falsa, la hipótesis de investigación (alternativa) debe ser cierta. Por las razones que se verán adelante, es más fácil mostrar apoyo de la hipótesis de investigación al presentar evidencia (datos muestrales) indicando que la hipótesis nula es falsa. Esto es, se busca evidencia en favor de la hipótesis de investigación mediante el uso de un método que resulta similar a una demostración por contradicción.

Aunque se desea obtener evidencia que apoye a la hipótesis alternativa (que se denotará por el símbolo H_a), es la hipótesis nula (denotada por el símbolo H_0) la hipótesis que será probada. Entonces H_0 especifica los valores de la hipótesis para uno o más parámetros de la población. Por ejemplo se puede estar interesado en probar la hipótesis nula de que la media de una población es igual a 50, con la esperanza de mostrar que en realidad ésta excede a 50. O puede estarse interesado en probar la hipótesis nula de que las medias de dos poblaciones, μ_1 y μ_2, son iguales, esperando demostrar que en realidad, por ejemplo, μ_1 es mayor que μ_2.

La decisión de aceptar o rechazar la hipótesis nula se basa en la información contenida en la muestra tomada de la población de interés. Los valores muestrales se usan para calcular un número que corresponde a un punto en la línea, el cual funciona como variable de decisión. A esta variable de decisión se le llama **estadística de prueba.** El conjunto de todos los posibles valores que la estadística de prueba puede tomar se divide en dos conjuntos, o regiones, uno que corresponde a la **región de rechazo** y el otro que corresponde a la **región de aceptación.** Si la estadística de prueba, al ser calculada a partir de una muestra en particular, toma un valor que se encuentra en la región de rechazo, entonces se rechaza la hipótesis nula y la hipótesis alternativa o de investigación es aceptada. Si la estadística de prueba toma un valor en la región de aceptación, entonces se acepta la hipótesis nula o bien se considera que no hubo evidencia para rechazarla. Adelante se explican las circunstancias que conducen a esta última decisión.

El procedimiento de decisión antes descrito está sujeto a dos tipos de error. Estos errores son inherentes a todo problema de decisión en donde se tengan dos selecciones posibles. Puede rechazarse la hipótesis nula H_0 cuando

en realidad es verdadera o puede aceptarse H_0 cuando es falsa y alguna alternativa cierta. Estos errores se conocen con el nombre de **error tipo I** y **error tipo II** respectivamente.

Definición

Un **error tipo I** en una prueba estadística es el error que se comete al rechazar la hipótesis nula cuando ésta es cierta. La probabilidad de cometer el error tipo I se denota por el símbolo α.

Un **error tipo II** en una prueba estadística es el error que se comete al no rechazar la hipótesis nula cuando ésta es falsa. La probabilidad de cometer el error tipo II se denota por el símbolo β.

Las dos posibilidades para la hipótesis nula, esto es ser falsa o verdadera, junto con las dos posibles decisiones del experimentador se presentan en la tabla de doble entrada, tabla 8.3. La ocurrencia de los errores tipo I y tipo II se señala en la tabla.

Tabla 8.3 *Tabla de decisión*

	HIPOTESIS NULA	
DECISION	CIERTA	FALSA
Rechazar H_0	error tipo I	decisión correcta
Aceptar H_0	decisión correcta	error tipo II

La bondad de una prueba estadística de hipótesis se evalúa mediante las probabilidades de cometer los errores tipo I y tipo II, que se denotan por los símbolos α y β respectivamente. Estas probabilidades, calculadas para las pruebas estadísticas elementales que se presentaron en el capítulo 6, ilustran la relación básica entre α, β y el tamaño de muestra n. Puesto que α es la probabilidad de que la estadística de prueba caiga en la región crítica (región de rechazo), suponiendo cierta H_0, **un incremento en el tamaño de la región crítica aumenta α,** y simultáneamente disminuye β, para un tamaño de muestra fijo. El reducir el tamaño de la región crítica disminuye α y aumenta β. Si se aumenta el tamaño de muestra entonces, se tiene más información en la cual basar la decisión y ambas α y β decrecerán.

La probabilidad β, de cometer el error de tipo II, varía dependiendo del verdadero valor del parámetro de la población. Por ejemplo suponga que se desea probar la hipótesis nula de que el parámetro binomial p es igual a $p_0 = .4$ (se usa un subíndice 0 para indicar el valor del parámetro especificado por H_0). Suponiendo que H_0 es falsa y que p es en realidad igual a un valor alternativo, digamos p_a. ¿Qué valor alternativo será detectado más fácilmente $p_a = .4001$ ó $p_a = 1.0$? Por supuesto si p es en realidad igual a 1.0, cada ensayo será un exito y los resultados muestrales producirán una fuerte evidencia en apoyo del rechazo de $H_0: p_0 = .4$. Por otra parte, $p_a = .4001$ está tan cercano a $p_0 = .4$ que será extremadamente difícil detectar p_a a menos de que se tenga una muestra muy grande. En otras palabras, la probabilidad β de aceptar H_0 varía dependiendo de la diferencia entre el verdadero valor de p y el valor de la hipótesis

nula p_0. La gráfica de la probabilidad β de error de tipo II como función del valor verdadero del parámetro se llama **curva característica de operación** para la prueba estadística. Nótese que las curvas características de operación para el problema de muestreo de aceptación en el capítulo 6 eran gráficas que expresaban β en función de p.

Puesto que la región de rechazo se especifica y se mantiene constante para una prueba dada, α también permanece constante y, como en el ejemplo de muestreo de aceptación, la curva característica de operación describe las características de la prueba estadística. Un aumento en el tamaño de la muestra, n, reduce β para todo valor del parámetro a probar. Entonces se tiene una curva característica de operación para cada tamaño de muestra. Esta propiedad de la curva característica de operación se ilustró en los ejercicios del capítulo 6.

Idealmente se tendrán algunos valores α y β que midan los riesgos de los respectivos errores que el experimentador está dispuesto a tolerar. También se tendrá en mente alguna desviación del valor de la hipótesis H_0, que se considera de importancia práctica y que se desea encontrar. La región de rechazo para la prueba se encontrará de acuerdo al valor α. Finalmente se selecciona un tamaño de muestra n para alcanzar un valor aceptable para β para la magnitud de desviación que se desea detectar. Esto puede hacerse mediante el uso de las curvas características de operación para distintos tamaños de muestra para la prueba seleccionada.

En la práctica, por lo común β es desconocido, ya sea porque nunca se calculó antes de realizar la prueba o porque puede resultar extremadamente difícil de calcular para la prueba. Entonces, en lugar de aceptar la hipótesis nula cuando la estadística de prueba cae en la región de aceptación debe de reservarse la decisión. Esto es, no debe aceptarse la hipótesis nula, a menos que se conozca el riesgo (medido por β) de tomar la decisión incorrecta. Nótese que esta situación de "no conclusión" no se presenta cuando la estadística de prueba cae en la región de rechazo. En este caso la hipótesis nula puede rechazarse (y la alternativa aceptarse) puesto que se conoce el valor de α, la probabilidad de rechazar la hipótesis nula cuando es cierta. El hecho de que β es a menudo desconocido explica el porqué se pretende apoyar la hipótesis alternativa por medio de un rechazo de la hipótesis nula. Cuando se toma esta decisión, la probabilidad α de que la decisión sea incorrecta es conocida.

Se verá en el texto como la hipótesis alternativa H_a ayuda en la localización de la región de rechazo.

8.14 Pruebas estadísticas con muestras grandes

Las pruebas de hipótesis usando muestras grandes, acerca de los parámetros μ, p, $(\mu_1 - \mu_2)$, y $(p_1 - p_2)$ se basan en estadísticas de prueba que siguen una distribución normal y por esta razón pueden considerarse como una misma prueba. Aquí, el razonamiento involucrado en la prueba, se presentará en forma muy general, haciendo referencia al parámetro de interés θ. Puede pen-

sarse, entonces, que θ representa a μ, $(\mu_1 - \mu_2)$, p o $(p_1 - p_2)$. La prueba específica para cada parámetro será ilustrada con ejemplos.

Suponga que se desea probar una hipótesis acerca del parámetro θ y que se tiene un estimador insesgado $\hat{\theta}$ el cual sigue una distribución normal con desviación estándar $\sigma_{\hat{\theta}}$. Si la hipótesis nula

$$H_0: \theta = \theta_0$$

es cierta, entonces la distribución muestral de $\hat{\theta}$ es normal con media θ_0, como se muestra en la figura 8.10.

Suponiendo que el objetivo del experimento es obtener evidencia para mostrar que θ es mayor que θ_0, la hipótesis alternativa es $H_a: \theta > \theta_0$, y se rechazará la hipótesis nula cuando $\hat{\theta}$ sea muy grande; "muy grande" significa, por supuesto, muchas desviaciones estándar $\sigma_{\hat{\theta}}$ alejado de θ_0. La región de rechazo se muestra en la figura 8.10. El valor de $\hat{\theta}$ que separa la región de rechazo y la región de aceptación, denotado por el símbolo C, se denomina el **valor crítico** de la estadística de prueba. La probabilidad de rechazo, suponiendo cierta la hipótesis nula H_0, es igual al área bajo la curva normal que queda sobre la región de rechazo. Esta área α aparece sombreada en la figura 8.10. Si se desea $\alpha = .05$, se rechaza la hipótesis nula cuando $\hat{\theta}$ está a más de $1.645\sigma_{\hat{\theta}}$ a la derecha de θ_0.

Definición

Una **prueba estadística de una sola cola** es aquella en la que la región de rechazo se localiza en solamente un extremo de la distribución muestral de la estadística de prueba. Para detectar $\theta > \theta_0$ la región de rechazo debe situarse en la cola (extremo) superior de la distribución de $\hat{\theta}$. Para detectar $\theta < \theta_0$ la región de rechazo debe situarse en la cola inferior de la distribución de $\hat{\theta}$.

Si se desea detectar discrepancias del tipo θ es mayor que θ_0 ó θ es menor que θ_0, la hipótesis alternativa es

$$H_a: \theta \neq \theta_0$$

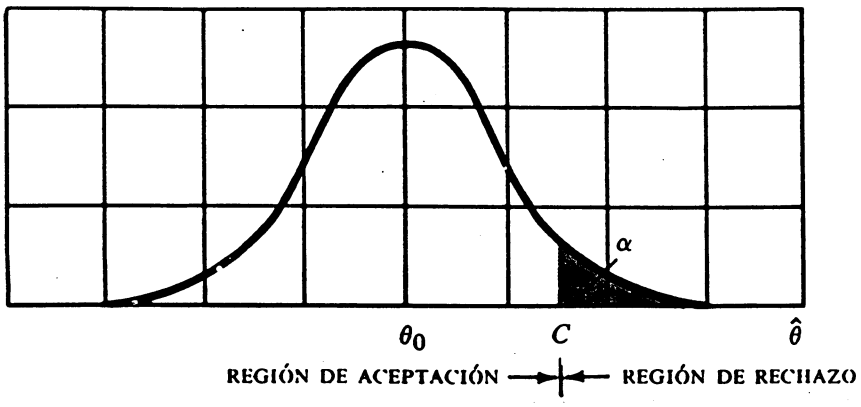


Figura 8.10 *Distribución de $\hat{\theta}$ cuando H_0 es cierta*

esto es

$$\theta > \theta_0 \qquad \text{ó} \qquad \theta < \theta_0$$

En este caso la probabilidad α de error de tipo I se reparte por igual entre las dos colas de la distribución normal, obteniéndose como resultado una **prueba estadística de dos colas.**

Definición

Una **prueba estadística de dos colas** es aquella que sitúa la región crítica en ambas colas de la distribución muestral de la estadística de prueba. Las pruebas de dos colas se usan para detectar ya sea cuando $\theta > \theta_0$ ó $\theta < \theta_0$.

El cálculo de β para pruebas estadísticas de una sola cola ha sido descrito anteriormente y puede facilitarse si se considera la figura 8.11. Cuando H_0 es falsa y $\theta = \theta_a$, la estadística de prueba $\hat{\theta}$ sigue una distribución normal con media θ_a en lugar de θ_0. La distribución de $\hat{\theta}$, suponiendo $\theta = \theta_a$, se ilustra en la figura mediante una curva normal sólida. La distribución de $\hat{\theta}$ bajo H_0 que se ilustra con una curva punteada, se usa para localizar la región de rechazo y el valor crítico C para la estadística de prueba $\hat{\theta}$. Puesto que β es la probabilidad de aceptar H_0 dado $\theta = \theta_a$, β es igual al área bajo la curva sólida localizada sobre la región de aceptación. Esta área, que aparece sombreada en la figura, puede ser calculada utilizando los métodos descritos en el capítulo 7.

Debe notarse que todos los estimadores discutidos en la sección precedente satisfacen los requerimientos de la prueba que se acaba de discutir cuando el tamaño de muestra es grande. Esto es, el tamaño de muestra debe ser suficientemente grande para que el estimador se distribuya aproximadamente normal de acuerdo al teorema central del límite y permita una estimación razonablemente buena de su desviación estándar. Es posible ahora probar hipótesis acerca de μ, p, $(\mu_1 - \mu_2)$ y $(p_1 - p_2)$.

La mecánica del procedimiento de prueba se simplifica al usar

$$z = \frac{\hat{\theta} - \theta_0}{\sigma_{\hat{\theta}}}$$

como estadística de prueba, como se ha notado ya en el ejemplo 7.11. Debe notarse que z es simplemente la desviación respecto a θ_0 de una variable alea-

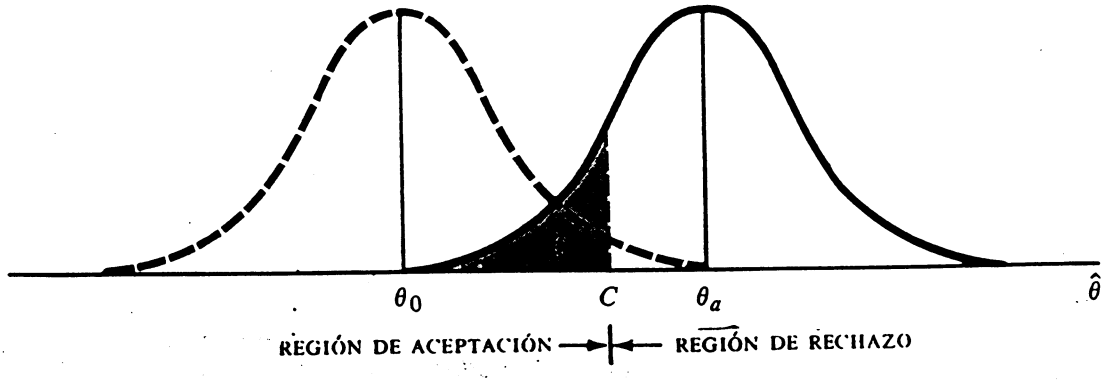


Figura 8.11 *Distribución de $\hat{\theta}$ cuando H_0 es falsa y $\theta = \theta_a$.*

toria normalmente distribuída, $\hat{\theta}$, expresada en unidades de $\sigma_{\hat{\theta}}$. Entonces para una prueba de dos colas con $\alpha = .05$, se debe rechazar H_0 cuando $z > -1.96$ ó $z < -1.96$, puesto que $P(z < -1.96 \text{ ó } z > 1.96) = .05$ cuando H_0 es cierta.

Como ha sido establecido previamente, el método de inferencia usado en una situación dada depende a menudo de la preferencia del experimentador. Algunas personas desean expresar su inferencia como un estimador, otras prefieren probar una hipótesis acerca del parámetro de interés.

El siguiente ejemplo ilustra el uso de la estadística z para probar la hipótesis de que el rendimiento promedio del producto químico es $\mu = 880$ tons por día contra la alternativa de que μ es mayor o menor que 880 tons por día. Ilustra al mismo tiempo la cercana relación existente entre una prueba de dos colas y los intervalos de confianza discutidos en secciones precedentes.

Ejemplo 8.10 Con los datos del ejemplo 8.1, pruebe la hipótesis de que el rendimiento promedio diario del producto químico es $\mu = 880$ tons contra la alternativa de que μ es o mayor o menor que 880 tons por día. La muestra basada en $n = 50$ observaciones produce una $\bar{y} = 871$ y $s = 21$ tons.

Solución Las hipótesis nula y alternativa son

$$H_0: \mu = 880 \qquad \text{y} \qquad H_a: \mu \neq 880$$

El estimador puntual para μ es $\bar{y}$. Por lo tanto la estadística de prueba es

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

Usando s en lugar de σ se tiene:

$$z = \frac{871 - 880}{21/\sqrt{50}} = -3.03$$

Para $\alpha = .05$ la región de rechazo es $z > 1.96$ ó $z < -1.96$. Puesto que el valor calculado de z, -3.03, cae en la zona de rechazo, se rechaza la hipótesis que $\mu = 880$ tons y se concluye que es menor. La probabilidad de rechazar siendo H_0 cierta es $\alpha = .05$. Por lo que se puede estar razonablemente seguro de que la decisión es correcta.

La prueba estadística basada en una estadística de prueba que sigue una distribución normal con un α dado y el intervalo de confianza del $(1 - \alpha)100\%$ de la sección 8.7 están relacionadas. El intervalo $(\bar{y} \pm 1.96\sigma/\sqrt{n})$ o aproximadamente (871 ± 5.82) para el ejemplo 8.10 está construido de manera tal que en muestreo sucesivo $(1 - \alpha)100\%$ de los intervalos contenga μ. Note que $\mu = 880$ tons no cac en el intervalo, lo que haría plausible rechazar $\mu = 880$ como un valor factible y concluir entonces que el rendimiento diario promedio es en realidad menor.

Otra similitud entre la prueba y el intervalo de confianza de la sección 8.7 es que la prueba es también aproximada puesto que se substituye σ con el valor aproximado s. Esto es, la probabilidad α de error tipo I seleccionada para la prueba no es precisamente .05. Es cercana a .05, suficientemente cer-

sana para todo propósito práctico, pero no es exacta. Lo mismo ocurre para todo procedimiento estadístico puesto que es imposible que todas y cada una de las suposiciones teóricas se cumplan exactamente.

El siguiente ejemplo ilustra el cálculo de β para la prueba estadística del ejemplo 8.10.

Ejemplo 8.11 Con referencia al ejemplo 8.10 calcule la probabilidad β de aceptar H_0 cuando en realidad μ es igual a 870 tons.

Solución La región de aceptación para la prueba del ejemplo 8.10 se localiza en el intervalo $(\mu_0 \pm 1.96\sigma_{\bar{y}})$. Si se substituyen los valores numéricos correspondientes se obtiene

$$880 \pm 1.96\frac{21}{\sqrt{50}} \qquad \text{ó} \qquad 874.18 \text{ a } 885.82$$

La probabilidad de aceptar H_0 dado $\mu = 870$, es igual al área bajo la distribución de frecuencias de la estadística de prueba $\bar{y}$ en el intervalo de 874.18 a 885.82. Puesto que $\bar{y}$ se distribuye normalmente con media 870 y $\sigma_{\bar{y}} = 21/\sqrt{50} = 2.97$, β es igual al área bajo esta curva normal localizada entre 874.18 y 885.82. Ver figura 8.12. Al calcular los valores de z correspondientes a 874.18 y 885.82 se tiene:

$$z_1 = \frac{\bar{y} - \mu}{\sigma/\sqrt{n}} = \frac{874.18 - 870}{21/\sqrt{50}} = 1.41$$

$$z_2 = \frac{\bar{y} - \mu}{\sigma/\sqrt{n}} = \frac{885.82 - 870}{21/\sqrt{50}} = 5.33$$

Entonces

$\beta = P(\text{aceptar } H_0 \text{ cuando } \mu = 870) = P(874.18 < \bar{y} < 885.82 \text{ cuando } \mu = 870) = P(1.41 < z < 5.33)$.

Como se observa en la figura 8.12 el área bajo la curva normal más allá de $\bar{y} = 885.82$ (ó $z = 5.33$) es casi nula. Por lo tanto

$$\beta = P(z > 1.41)$$

De la tabla 3 en el apéndice, se encuentra que el área entre $z = 0$ y $z = 1.41$ es de .4207. Por lo tanto $\beta = .5 - .4207 = .0793$.

Entonces la probabilidad de aceptar H_0, dado que en realidad $\mu = 870$, es .0793 o aproximadamente 8 de 100.

Si se calcula μ para un valor de μ_a más cercano a $\mu_0 = 880$, la distribución que se muestra en la figura 8.12 se desplazaría a la derecha y β aumentaría. Por el contrario entre mayor la distancia entre μ_a y μ_0, menor será el valor de β. Esto es, se cometen menos errores en decidir que la media es diferente de μ_0 (H_0 es falsa) cuando la diferencia entre μ_0 y μ_a es grande.

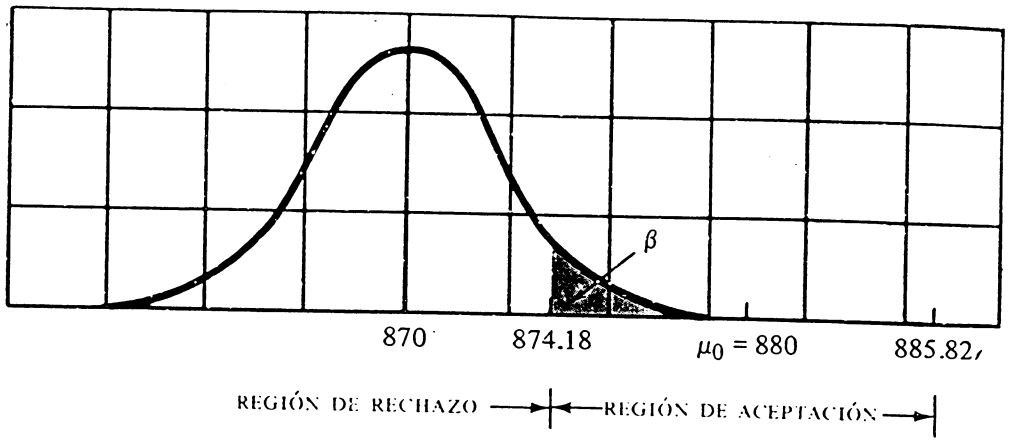


Figura 8.12 *Cálculo de β en el ejemplo 8.11*

Ejemplo 8.12 Se sabe que aproximadamente 1 de 10 fumadores prefiere la marca de cigarillos A. Después de una campaña publicitaria en una región, se entrevistó a 200 fumadores para determinar la eficiencia de la campaña. El resultado de esta muestra es que un total de 26 personas indicaron preferencia por la marca A. ¿Puede considerarse que estos datos presentan evidencia suficiente para indicar un aumento en la aceptación de la marca A? (Note que para propósitos prácticos este problema es idéntico al de la margarina del ejemplo 7.10.)

Solución Puede suponerse que la muestra satisface los requerimientos de un experimento binomial. La cuestión puede plantearse como la prueba de hipótesis

$$H_0: p = .10 \text{ (el programa resulta ineficiente)}$$

contra la alternativa

$$H_a: p > .10 \text{ (el programa es efectivo)}$$

Se utilizará una prueba de una sola cola puesto que se desea principalmente detectar un valor de p mayor que .10. Para este caso puede demostrarse que la probabilidad β de error de tipo II se minimiza al situar toda la región de rechazo en la cola superior de la distribución de la estadística de prueba.

El estimador de p es $\hat{p} = y/n$ y la estadística de prueba es

$$z = \frac{\hat{p} - p_0}{\sigma_{\hat{p}}} = \frac{\hat{p} - p_0}{\sqrt{p_0 q_0 / n}}$$

(*Nota:* Esta estadística de prueba es algebraicamente equivalente a:

$$z = \frac{y - np_0}{\sqrt{np_0 q_0}}$$

la estadística de prueba usada en el ejemplo 7.10.)

Una vez más, se requiere de un valor de p para calcular el valor $\sigma_{\hat{p}} = \sqrt{pq/n}$ que aparece en el denominador de z. Puesto que el valor hipo-

tético que se supone para p es $p = p_0$ parecería razonable usar p_0 como una aproximación de p. Note que esta aproximación difiere de la usada en el problema de estimación donde al no conocer p se seleccionó $\hat{p}$ como la mejor aproximación. Esta aparente inconsistencia tendrá poco efecto sobre la inferencia, sea ésta una estimación o una prueba, cuando n es grande.

Al escoger $\alpha = .05$, se rechaza H_0 cuando $z > 1.645$ y al substituir los valores numéricos en la expresión para la estadística de prueba se tiene

$$z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0 q_0}{n}}} = \frac{.13 - .10}{\sqrt{\frac{(.10)(.90)}{200}}} = 1.41$$

El valor calculado de $z = 1.41$ no cae en la región de rechazo y por lo tanto no se rechaza H_0, la hipótesis de que la proporción de fumadores que prefiere la marca A es .10. No hay evidencia suficiente para indicar que la campaña publicitaria ha sido efectiva y que el porcentaje de fumadores de la marca A ha aumentado como resultado del esfuerzo publicitario.

¿Debe aceptarse H_0? No, hasta que haya sido establecido un valor alternativo de p mayor que $p_0 = .10$ y que se considere de significancia práctica y el valor de β haya sido calculado para esa alternativa particular. Entonces, si el valor de β es suficientemente pequeño, puede aceptarse H_0, conociendo cual es el riesgo de una decisión equivocada.

Los ejemplos 8.10 y 8.12 ilustran un aspecto relevante. Si los datos ofrecen suficiente evidencia para rechazar H_0 la probabilidad α de una decisión incorrecta es conocido de antemano, puesto que α se usa para determinar la región de rechazo. Como α es por lo común pequeña se puede estar razonablemente seguro de haber hecho una decisión correcta. Por otra parte, si los datos no muestran evidencia suficiente para rechazar H_0, la conclusión no es tan obvia. No necesariamente se sigue que $p = p_0$ cuando no se rechaza la hipótesis nula. Más bien se concluye que la evidencia es insuficiente para indicar que p es mayor que p_0. En el ejemplo 8.12, a un nivel de significancia $\alpha = .05$, no se hubiera rechazado la hipótesis para cualquier valor hipotético de p mayor que $p_0 = .095$. No hay evidencia para afirmar que p es mayor que .095, .097, .10, .13 o cualquier otro valor mayor que .095. Sin embargo, sólo uno de estos valores, $p = .10$, tiene relevancia en el contexto del problema. Idealmente, al seguir el procedimiento estadístico delineado en la sección 8.13, se tiene el valor alternativo p_a de significancia práctica previamente determinado y se ha escogido n de manera tal que β sea pequeño. Desafortunadamente muchos experimentadores no proceden en esta forma ideal. Alguien escoge el tamaño de muestra y al experimentador o al estadístico se le pide que evalúe la evidencia.

El cálculo de β no es muy difícil para los procedimientos de prueba estadísticos delineados en esta sección, pero puede ser extremadamente difícil, incluso más allá de la capacidad del principiante, en otras situaciones. Resulta mucho más simple *no rechazar* H_0 en lugar de aceptarla; después *estimar* p usando un intervalo de confianza, que dará un rango o conjunto de valores factibles para p.

Ejemplo 8.13 Buscando reducir el gasto anual de gasolina de 350 millones de galones, el Servicio Postal de los Estados Unidos ha adquirido algunas vagonetas eléctricas para el correo a manera experimental. Los expertos suponen que se producirá un ahorro si se usan las vagonetas eléctricas en las rutas planas y se hace uso de aquellas horas con tarifa eléctrica reducida para recargar las unidades. Se hace un experimento para comparar los costos de operación en el que se usan $n_1 = 100$ vagonetas de gasolina y $n_2 = 100$ vagonetas eléctricas. El costo por milla para operar los vehículos de gasolina es en promedio $\bar{y} = 6.70$ centavos por milla con una varianza de $s^2 = .36$ mientras que para las vagonetas eléctricas se observó una media y una varianza del costo de operación por milla de $\bar{y} = 6.54$ y $s^2 = .40$ respectivamente. ¿Muestran estos datos evidencia suficiente para indicar una diferencia significativa en el costo promedio de operación entre los vehículos convencionales de gasolina y las vagonetas eléctricas?

Solución Se desea probar la hipótesis de que la diferencia $(\mu_1 - \mu_2)$ entre las dos medias de la población es igual a algún valor preespecificado (D_0). Para el ejemplo se usará $D_0 = 0$, esto es que no hay diferencia en los costos promedio de operación para los dos tipos de vehículos.

Como se recordará $(\bar{y}_1 - \bar{y}_2)$ es un estimador insesgado para $(\mu_1 - \mu_2)$ que se distribuye aproximadamente normal cuando n_1 y n_2 son grandes. Además se sabe que la desviación estándar de $(\bar{y}_1 - \bar{y}_2)$ es

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

Entonces

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

sirve como estadística de prueba cuando σ_1^2 y σ_2^2 se conocen o s_1^2 y s_2^2 dan una buena estimación para σ_1^2 y σ_2^2 (esto es cuando n_1 y n_2 son mayores que 30).

Para este ejemplo, se tiene

$$H_0: \mu_1 - \mu_2 = D_0 = 0$$
$$H_a: \mu_1 - \mu_2 \neq 0$$

Al substituir los valores numéricos en la expresión para la estadística de prueba se obtiene

$$z = \frac{6.70 - 6.54}{\sqrt{\frac{.36}{100} + \frac{.40}{100}}} \approx 1.83$$

y usando una prueba de dos colas con $\alpha = .05$ se rechaza H_0 cuando $z > 1.96$ o $z < -1.96$. Puesto que z no cae en la región de rechazo no se rechaza la hipótesis nula.

Nótese, sin embargo, que si se hubiera escogido $\alpha = .10$ la región de rechazo hubiera sido $z > 1.645$ o $z < -1.645$ y la hipótesis nula se hubiera rechazado.

La decisión de aceptar o rechazar depende del riesgo que se desea tolerar. En el ejemplo 8.13 si se escoge $\alpha = .05$, la hipótesis nula no se rechaza, pero no se debe aceptar H_0 ($\mu_1 = \mu_2$) sin investigar la probabilidad de error de tipo II.

Por otra parte si α se escogiera igual a .10 la hipótesis nula se hubiera rechazado. Si no se contara con más información, podría pensarse en rechazar la hipótesis nula de no diferencia en los costos promedio de operación para los vehículos de gasolina y para los eléctricos. La probabilidad de rechazar H_0 cuando ésta es cierta es solamente $\alpha = .10$ y puede pensarse que se ha tomado una decisión razonablemente buena.

Ejemplo 8.14 Los administradores de los hospitales en muchos casos se encargan de obtener y calcular algunas estadísticas que son de suma importancia para los médicos y para los encargados de decidir en el hospital. En los registros de un hospital se tiene que 52 hombres en una muestra de 1000 hombres y 23 mujeres en una muestra de 1000 mujeres ingresaron al hospital a causa de alguna enfermedad cardíaca. ¿Puede o no considerarse que estos datos presentan evidencia suficiente en el sentido de que existe una mayor tasa de afecciones cardiacas en los hombres que ingresan al hospital?

Solución Puede suponerse que el número de pacientes que ingresa a un hospital a causa de alguna afección cardiaca sigue aproximadamente una distribución binomial para ambos hombres y mujeres, con parámetros p_1 y p_2 respectivamente. Puede establecerse que se desea probar la hipótesis de que existe una diferencia entre p_1 y p_2, $p_1 - p_2 = D_0$. (Por ejemplo, en este caso se desea probar la hipótesis $D_0 = 0$.) Se recordará que para muestras grandes, el estimador $\hat{p}_1 - \hat{p}_2$ sigue una distribución aproximadamente normal con media $(p_1 - p_2)$ y con desviación estándar

$$\sigma_{(\hat{p}_1 - \hat{p}_2)} = \sqrt{\frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}}$$

Entonces

$$z = \frac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)}{\sigma_{(\hat{p}_1 - \hat{p}_2)}}$$

sigue una distribución normal estandarizada. De aquí que z pueda usarse como estadística de prueba para probar

$$H_0: (p_1 - p_2) = D_0$$

Ahora, se deben tener aproximaciones adecuadas para p_1 y p_2 para calcular $\sigma_{(\hat{p}_1 - \hat{p}_2)}$. Para estas aproximaciones pueden considerarse los dos casos siguientes.

Supone que p_1 es igual a p_2, es decir

$$H_0: p_1 = p_2 \quad \text{o} \quad (p_1 - p_2) = 0$$

entonces $p_1 = p_2 = p$ y el mejor estimador de p se obtiene al combinar los datos de ambas muestras. Si y_1 y y_2 son los números de éxitos que se obtuvieron en las muestras, entonces:

$$\hat{p} = \frac{y_1 + y_2}{n_1 + n_2}$$

y entonces la estadística de prueba resulta

$$z = \frac{(\hat{p}_1 - \hat{p}_2) - 0}{\sqrt{\frac{\hat{p}\hat{q}}{n_1} + \frac{\hat{p}\hat{q}}{n_2}}} = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}\hat{q}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

Caso 2. Si por otra parte se supone que D_0 no es igual a cero, esto es

$$H_0: (p_1 - p_2) = D_0$$

con $D_0 \neq 0$, entonces los mejores estimadores para p_1 y p_2 son $\hat{p}_1$ y $\hat{p}_2$ respectivamente y la estadística de prueba resulta:

$$z = \frac{(\hat{p}_1 - \hat{p}_2) - D_0}{\sqrt{\frac{\hat{p}_1\hat{q}_1}{n_1} + \frac{\hat{p}_2\hat{q}_2}{n_2}}}$$

Para muchos problemas prácticos relacionados con la comparación de dos poblaciones binomiales el investigador desea probar la hipótesis de que $(p_1 - p_2) = D_0 = 0$. En este ejemplo se desea probar

$$H_0: (p_1 - p_2) = 0$$

contra la alternativa

$$H_a: (p_1 - p_2) > 0$$

Note que se usará una prueba de una sola cola puesto que si existe una diferencia, se desea detectar si $p_1 > p_2$, por lo que se rechazará la hipótesis nula en favor de la alternativa si $(\hat{p}_1 - \hat{p}_2)$ es grande o, equivalentemente, si el valor calculado de z es grande. De hecho, si se escoge $\alpha = .05$ se rechazará H_0 cuando $z > 1.645$.

El estimador combinado para p que se necesita para calcular $\sigma_{(\hat{p}_1 - \hat{p}_2)}$ es

$$\hat{p} = \frac{y_1 + y_2}{n_1 + n_2} = \frac{52 + 23}{1000 + 1000} = .0375$$

La estadística de prueba es

$$z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}\hat{q}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} = \frac{.052 - .023}{\sqrt{(.0375)(.9625)\left(\frac{1}{1000} + \frac{1}{1000}\right)}} = 3.41$$

Puesto que el valor de z cae en la región de rechazo, se rechaza la hipótesis de que $p_1 = p_2$ y se concluye que los datos presentan evidencia suficiente de

que el porcentaje de hombres que ingresan al hospital a causa de una afección cardíaca es más alto que el de las mujeres. Note que este resultado no establece que la tasa de afección cardíaca es más alto para hombres. Es posible que sean menos las mujeres que ingresan al hospital cuando padecen tal enfermedad.

Ejercicios

8.24. Un fabricante afirma que al menos 95% del equipo que ha surtido para cierta fábrica cumple con las especificaciones. Se examina una muestra de 700 piezas de equipo y se encuentra que 53 de ellas son defectuosas ¿Puede decirse que los datos proporcionan suficiente evidencia para rechazar la afirmación del fabricante? Use $\alpha = .05$.

a. Establezca la hipótesis nula que se ha de probar.
b. Establezca la hipótesis alternativa.
c. Proceda a realizar la prueba estadística y presente sus conclusiones.

8.25. En vista de la disminución de recursos energéticos, la Administración Nacional de Aeronáutica y del Espacio (NASA) de los Estados Unidos se ha dado a la tarea de encontrar sitios en aquel país en donde resulte factible instalar molinos de viento para generar energía eléctrica. Un oficial de la NASA, D.J. Vargo, ha dicho que la velocidad del viento debe promediar al menos 15 millas por hora para que un sitio pueda considerarse aceptable.* Se hicieron 36 mediciones de la velocidad del viento a intervalos aleatorias en un sitio bajo consideración para instalar un molino; la velocidad del viento promedió $\bar{y} = 14.2$ mph. con desviación estándar de 3 mph. ¿Puede considerarse que los datos indican que el sitio no satisface los requerimientos de la NASA para la instalación de un generador de energía a base de viento? Use $\alpha = .10$.

8.26. Continuando con el ejercicio 8.25 discuta el significado de los errores tipo I y tipo II en el problema de la selección de sitios. Si usted fuese un oficial de la NASA, ¿recomendaría usted el uso de un valor extremadamente pequeño para α (.01) o bien un valor moderado (.05) o un valor razonablemente grande (.10) para determinar la aceptabilidad de un sitio en base a la velocidad promedio del viento? Explique porqué.

*"Wind Power: Still Becalmed," *Environmental News*, julio de 1975.

8.27. Se estudia el rendimiento de dos tipos de valores, el valor tipo A y el valor tipo B. El rendimiento de estos sólo se conoce hasta después de la venta. Se tienen registradas 30 tasas de rendimiento anterior para valores tipo A y 36 para valores tipo B. De estas muestras se obtuvieron las estadísticas en la tabla siguiente.

VALORES TIPO A	VALORES TIPO B
$\bar{y}_1 = 7.91\%$	$\bar{y}_2 = 8.12\%$
$s_1^2 = 1.85$	$s_2^2 = 1.98$

¿Presentan estos datos suficiente evidencia de que existe una diferencia en el rendimiento de ambos valores? Use $\alpha = .10$.

8.28. En un estudio para averiguar los efectos de usar modelos femeninos en la publicidad para automóviles, a un grupo de 50 hombres, el grupo A, se le mostró la fotografía de un automóvil con una modelo femenina y la de otro automóvil del mismo precio pero sin modelo. A otro grupo, el grupo B, de 50 hombres se les mostraron ambos automóviles sin modelo femenina. En el grupo A el automóvil que aparecía con la modelo fue considerado más lujoso por 37 de los entrevistados, en el grupo B el mismo automóvil fue juzgado como más lujoso por 28 de los entrevistados. ¿Se considera que estos datos indican que el usar una modelo femenina influye en el lujo aparente de un automóvil? Use una prueba de una sola cola con $\alpha = .05$

8.15 Algunos comentarios sobre la teoría de pruebas de hipótesis

Como se ha delineado en la sección 8.13, la teoría de la prueba estadística de una hipótesis es un procedimiento bien definido que permite al investigador

ya sea aceptar o rechazar la hipótesis nula con riesgos de error medidos por las probabilidades α y β. Desafortunadamente, como se ha notado, este marco teórico no es suficiente para todas las situaciones prácticas.

La teoría requiere que se tenga habilidad para especificar una alternativa significativa que permita el cálculo de β, la probabilidad de error tipo II, para valores alternativos del parámetro. Esto puede llevarse a cabo para muchas pruebas, incluyendo la prueba discutida en la sección 8.13, aunque el cálculo de β para diversas alternativas y para distintos tamaños de muestra puede, en algunos casos, resultar una tarea formidable. Por otra parte, en algunos casos puede resultar extremadamente difícil especificar alternativas para H_0 que resulten de significancia *práctica*. Esto puede ocurrir cuando se desea probar hipótesis acerca de valores de conjuntos de parámetros, situación que se encontrará en el capítulo 17, al analizar datos clasificados.

Este obstáculo no invalida el uso de las pruebas estadísticas, sino que sugiere que se tenga precaución para llegar a una conclusión cuando la evidencia es insuficiente para rechazar la hipótesis nula. La dificultad para especificar alternativas significativas a la hipótesis nula, junto con la dificultad que puede encontrarse para calcular y tabular β para casos más complicados que los aquí presentados, justifica que su discusión sea evadida en un texto de nivel introductorio. De aquí que se acuerde adoptar el procedimiento descrito en el ejemplo 8.11 cuando no se dispone de valores tabulados de β para la prueba en cuestión. Cuando la estadística de prueba cae en la región de aceptación, la decisión será "no rechazar" en lugar de "aceptar" la hipótesis nula. Pueden obtenerse conclusiones adicionales al calcular un estimador por intervalo para el parámetro o mediante la consulta de uno de los varios manuales estadísticos para valores tabulados de β. No debe ocasionar sorpresa el hecho de que valores tabulados de β resulten inaccesibles para algunas de las pruebas estadísticas más complicadas.

La probabilidad α de cometer un error tipo I se denomina a menudo **nivel de significancia** de la prueba. Este término se originó de la manera siguiente. La probabilidad del valor observado de la estadística o algún valor aún más contradictorio a la hipótesis nula, mide en algún sentido el peso de la evidencia en favor de la hipótesis alternativa. Algunos experimentadores reportan los resultados de una prueba como significante (rechazo) al 5% pero no al 1%. Esto significa que se rechazaría H_0 si α fuese .05 pero no si α fuera .01. Este modo de pensar no contradice al procedimiento de escoger de antemano la prueba a la recolección de datos, sino presenta un modo conveniente de publicar los resultados estadísticos para una investigación y le permite al lector escoger α y β como lo desee.

Finalmente debe comentarse sobre la selección de una prueba de una o dos colas para una situación dada. Debe enfatizarse que esta decisión está dada por los aspectos prácticos del problema y dependerá del valor alternativo del parámetro que el experimentador está deseando detectar. Si se tuviera que soportar una pérdida financiera seria si $\theta > \theta_0$ pero no si $\theta < \theta_0$ se concentraría la atención en valores de θ mayores que θ_0 y por lo tanto se rechazaría usando sólo la cola superior de la distribución de la estadística de prueba. Por otra parte si se está igualmente interesado en detectar valores mayores o menores que θ_0 se usaría una prueba de dos colas.

8.16 Resumen

El material del capítulo 8 ha estado dirigido hacia dos objetivos. Primero, se han discutido los varios métodos de inferencia, así como los procedimientos para evaluar la bondad de estos. Se han presentado algunos métodos de estimación y de prueba de hipótesis que, merced al teorema central del límite, se basan en los resultados del capítulo 7. Las técnicas resultantes son de gran utilidad práctica y al mismo tiempo ilustran los principios básicos de la inferencia estadística.

Es posible hacer inferencias acerca de los parámetros de una población ya sea mediante una estimación o al probar una hipótesis acerca de su valor. Un parámetro puede estimarse por medio de un estimador puntual o por medio de un intervalo. La medida de bondad de un estimador puntual está dada por la cota del error de estimación. Similarmente el coeficiente de confianza y la longitud del intervalo miden la bondad de un intervalo de confianza.

La prueba estadística de una hipótesis acerca del parámetro o los parámetros de una población conduce idealmente a su aceptación o rechazo. Prácticamente se llega a una decisión de rechazo o no rechazo. Las probabilidades de tomar decisiones incorrectas, esto es de cometer los errores tipo I y tipo II, miden la bondad del procedimiento de decisión. Mientras que una prueba de hipótesis puede resultar adecuada para situaciones concretas, (por ejemplo el muestreo de aceptación) parece que la estimación sería la meta eventual de muchos investigadores experimentales. De aquí que resulte deseable el tener opción al seleccionar un método de inferencia.

Todos los intervalos de confianza y las pruebas de hipótesis que se describen en este capítulo están basados en el teorema central del límite y son por lo tanto aplicables a muestras grandes. Cuando n es grande, cada uno de los estimadores respectivos y estadísticas de prueba poseen para todo propósito práctico una distribución normal. Este resultado junto con las propiedades de la distribución normal estudiadas en el capítulo 7, permite la construcción de intervalos de confianza y el cálculo de α y β para las pruebas estadísticas.

Ejercicios Complementarios

8.29. Defina y discuta los siguientes conceptos y el papel que desempeñan en la inferencia estadística:
a. errores tipo I y tipo II.
b. las probabilidades α y β.
c. el nivel de significancia.
d. intervalo de confianza.
e. la prueba z.

8.30. Explique el papel del teorema central del límite en los temas presentados en este capítulo.

8.31. Explique la diferencia entre una prueba de hipótesis de una cola y una de dos colas.

8.32. Para cada una de las siguientes situaciones, sugiera si es apropiado el uso de una prueba de una cola o es más indicada una prueba de dos colas. Para cada caso justifique su razonamiento y escriba entonces la hipótesis nula y la hipótesis alternativa.

a. El estudio de un economista para determinar si el nivel de desempleo ha cambiado significativamente durante el año pasado.

b. Un estudio para verificar la suposición del gerente de crédito que afirma que el balance promedio por cuenta de los clientes que usan crédito en una tienda es de al menos $50.

c. Una investigación de mercado para determinar si

HARNETT, D. L., y J. L. MURPHY. *Introductory Statistical Analysis*. Reading, Mass.: Addison-Wesley, 1975. Capítulos 6 y 7.

HOEL, P. G., y R. J. JESSEN. *Basic Statistics for Business and Economics*. New York: Wiley, 1971. Capítulos 6 y 7.

SUMMERS, G. W., y W. S. PETERS. *Basic Statistics in Business and Economics*. Belmont, Calif.: Wadsworth, 1973. Capítulos 9 y 10.