SAH

Introducción a la econometría

Un enfoque moderno

APENDICE C

4a. edición

Jeffrey M. Wooldridge
Michigan State University

Traducción
Ma. del Carmen Enriqueta Hano Roa
Érika M. Jasso Hernan D´Borneville
Traductoras profesionales

Revisión técnica
Roberto Palma Pacheco
Centro de Alta Dirección en Economía y Negocios
Universidad Anáhuac

Domingo Rodríguez Benavides
Facultad de Economía
Universidad Nacional Autónoma de México

319855

CENGAGE Learning

Australia • Brasil • Corea • España • Estados Unidos • Japón • México • Reino Unido • Singapur

Apéndice C

Fundamentos de estadística matemática

C.1 Poblaciones, parámetros y muestreo aleatorio

La inferencia estadística involucra saber algo de una población, dada la disponibilidad de una muestra de esa población. Por **población** se entiende cualquier grupo bien definido de sujetos, que podrían ser individuos, empresas, ciudades o muchas otras posibilidades. "Saber" puede significar muchas cosas, las cuales se dividen en categorías de *estimación* y *prueba de hipótesis*.

Un par de ejemplos lo ayudarán a comprender estos términos. En la población de todos los adultos que trabajan en Estados Unidos, los economistas laborales están interesados en conocer el rendimiento de la educación, medido como el incremento porcentual promedio en los ingresos, dado otro año de educación. Sería poco práctico y costoso obtener la información sobre los ingresos y la educación de toda la población laboral en Estados Unidos, pero es posible obtener datos de un subconjunto de la población. Mediante los datos recabados, un economista laboral puede reportar que su mejor estimación del rendimiento por otro año de educación es de 7.5%. Esto es un ejemplo de *estimación puntual*. O quizá reporte un rango, como el "rendimiento de la educación se halla entre 5.6 y 9.4%." Éste es un ejemplo de *estimación de intervalo*.

Un economista urbano tal vez quiera saber si los programas vecinales de vigilancia contra la delincuencia están asociados con menores tasas de delitos. Después de comparar tales tasas en los vecindarios con y sin tales programas en una muestra de la población, el economista podría formular una de dos conclusiones: los programas de vigilancia vecinal afectan la delincuencia, o no lo hacen. Este ejemplo cae bajo el rubro de la prueba de hipótesis.

El primer paso en la inferencia estadística es identificar a la población de interés. Esto parece obvio, pero es importante ser muy específico. Una vez que se ha identificado a la población, se puede especificar un modelo para la relación poblacional de interés. Tales modelos involucran las distribuciones de probabilidad o las características de distribuciones de probabilidad, y éstas dependen de parámetros desconocidos. Los parámetros simplemente son constantes que determinan las direcciones y fortalezas de las relaciones entre las variables. En el ejemplo anterior de la economía laboral, el parámetro de interés es el rendimiento de la educación sobre la población.

Muestreo

Para analizar la inferencia estadística, se utilizará el escenario más simple posible. Sea Y una variable aleatoria que representa a una población con una función de densidad de probabilidad $f(y;\theta)$, que depende de un sólo parámetro θ. Se supone que se conoce la función de densidad de probabilidad (fdp) de Y, salvo por el valor de θ; diferentes valores de θ implican diferentes distribuciones poblacionales y, por tanto, lo que interesa aquí es el valor de θ. Es posible obtener ciertos tipos de muestras de la población, entonces podemos conocer algo acerca de θ. El esquema de muestreo más sencillo de manejar es el muestreo aleatorio.

Muestreo aleatorio. Si $Y_1, Y_2, \ldots, Y_n$ son variables aleatorias independientes con una función de densidad de probabilidad común $f(y;\theta)$, entonces se dice que $\{Y_1, \ldots, Y_n\}$ es una **muestra aleatoria** de $f(y;\theta)$ [o muestra aleatoria de la población representada por $f(y;\theta)$].

Cuando $\{Y_1, \ldots, Y_n\}$ es una muestra aleatoria de la densidad $f(y;\theta)$, también se dice que Y_i son *variables aleatorias independientes, idénticamente distribuidas* (o *i.i.d.*) de $f(y;\theta)$. En algunos casos no se necesitará especificar por completo cuál es la distribución común.

La naturaleza aleatoria de $Y_1, Y_2, \ldots, Y_n$ en la definición de muestreo aleatorio refleja el hecho de que muchos resultados diferentes son posibles antes de que se realice realmente la muestra. Por ejemplo, si se obtiene el ingreso familiar para una muestra de $n = 100$ familias en Estados Unidos, los ingresos que se observan, por lo general, difieren para cada muestra diferente de 100 familias. Una vez que se obtenga la muestra se tiene un conjunto de números, por ejemplo, $\{y_1, y_2, \ldots, y_n\}$, que constituyen los datos con los que se trabajará. Si es o no correcto suponer que la muestra proviene de un esquema de muestreo aleatorio requiere conocer algo del proceso real de muestreo.

Las muestras aleatorias de una distribución de Bernoulli suelen utilizarse para ilustrar conceptos estadísticos, y también surgen en aplicaciones empíricas. Si $Y_1, Y_2, \ldots, Y_n$ son variables independientes aleatorias y cada una se distribuye como Bernoulli(θ), de manera que $P(Y_i = 1) = \theta$ y $P(Y_i = 0) = 1 - \theta$, entonces $\{Y_1, Y_2, \ldots, Y_n\}$ constituye una muestra aleatoria de la distribución de Bernoulli(θ). Como ilustración, considere el ejemplo de las reservaciones de la aerolínea que se realizó en el apéndice B. Cada Y_i denota si el cliente i muestra su reservación; $Y_i = 1$ si el pasajero i se presenta y $Y_i = 0$ de otra forma. Aquí, θ es la probabilidad de que una persona elegida al azar de la población de todas las personas que hicieron reservaciones aéreas se presente para su reservación.

En muchas otras aplicaciones se puede suponer que las muestras aleatorias se extraen de una distribución normal. Si $\{Y_1, \ldots, Y_n\}$ es una muestra aleatoria de la población Normal(μ,σ^2) entonces la población está caracterizada por dos parámetros, la media μ y la varianza σ^2. El interés principal suele enfocarse en μ, pero σ^2 es de interés por derecho propio debido a que las inferencias relativas a μ suelen requerir conocimientos acerca de σ^2.

C.2 Propiedades de muestras finitas de los estimadores

En esta sección, se estudia lo que se conoce como propiedades de los estimadores en muestras finitas. El término "muestra finita" proviene del hecho que las propiedades son válidas para una de cualquier tamaño, sin importar si es grande o pequeña. En ocasiones, éstas reciben el nombre de propiedades de muestras pequeñas. En la sección C.3, se cubrirán las "propiedades asintóticas", que tienen que ver con el comportamiento de los estimadores a medida que el tamaño de la muestra crece sin límite.

Estimadores y estimaciones

Para estudiar las propiedades de los estimadores se debe definir lo que se entiende por estimador. Dada una muestra aleatoria $\{Y_1, Y_2, \ldots, Y_n\}$ extraída de una distribución de la población que depende de un parámetro desconocido θ, un **estimador** de θ es una regla que asigna a cada resultado posible de la muestra un valor de θ. La regla se especifica antes de realizar cualquier muestreo y ésta es la misma, sin considerar realmente los datos obtenidos.

Como un ejemplo de un estimador, sea $\{Y_1, \ldots, Y_n\}$ una muestra aleatoria de una población con media μ. Un estimador natural de μ es el promedio de la muestra aleatoria:

$$\bar{Y} = n^{-1}\sum_{i=1}^{n} Y_i. \tag{C.1}$$

$\bar{Y}$ recibe el nombre de **promedio muestral**, pero a diferencia del apéndice A, donde se definió el promedio muestral de un conjunto de números como un estadístico descriptivo, $\bar{Y}$ ahora se considera como un estimador. Dado cualquier resultado de las variables aleatorias $Y_1, \ldots, Y_n$, se utiliza la misma regla para estimar μ: simplemente se promedian. Para los datos reales $\{y_1, \ldots, y_n\}$, la **estimación** es sólo el promedio en la muestra: $\bar{y} = (y_1 + y_2 + \ldots + y_n)/n$.

Ejemplo C.1

[Tasas de desempleo en las ciudades]

Suponga que se obtiene la siguiente muestra de las tasas de desempleo en 10 ciudades de Estados Unidos:

Ciudad	Tasa de desempleo
1	5.1
2	6.4
3	9.2
4	4.1
5	7.5
6	8.3
7	2.6
8	3.5
9	5.8
10	7.5

La estimación de la tasa de desempleo promedio en Estados Unidos es $\bar{y} = 6.0$. Cada muestra, por lo general, tiene como resultado una estimación diferente. Pero la *regla* para obtener la estimación es la misma, sin importar qué ciudades aparecen en la muestra, o cuántas.

En términos más generales, un estimador W de un parámetro θ se puede expresar como una fórmula matemática abstracta:

$$W = h(Y_1, Y_2, \ldots, Y_n), \tag{C.2}$$

para alguna función conocida h de las variables aleatorias $Y_1, Y_2, \ldots, Y_n$. Como en el caso especial del promedio muestral, W es una variable aleatoria pues depende de la muestra aleatoria: conforme se obtienen diferentes muestras aleatorias de la población, el valor de W puede cambiar. Cuando un conjunto de números en particular, por ejemplo $\{y_1, y_2, \ldots, y_n\}$, se inserta a la función h, se obtiene una *estimación* de θ, denotada como $w = h(y_1, \ldots, y_n)$. En ocasiones, W es llamado un estimador puntual y w una estimación puntual para distinguirlos de los estimadores y estimaciones de *intervalo*, que se estudiarán en la sección C.5.

Para evaluar los procedimientos de estimación, se estudian las diferentes propiedades de la distribución probabilística de la variable aleatoria W. La distribución de un estimador suele recibir el nombre de **distribución de muestreo**, debido a que ésta describe la probabilidad de varios resultados de W a través de diferentes muestras aleatorias. Debido a que existen reglas ilimitadas en la combinación de datos para estimar parámetros, se requieren algunos criterios sensibles para elegir entre los estimadores, o al menos para no considerar algunos de ellos. Por tanto, se abandona el ámbito de la estadística descriptiva, donde se calculan cuestiones como el promedio muestral para resumir simplemente un cuerpo de datos. En la estadística matemática se estudian las distribuciones muestrales de los estimadores.

Insesgadez

En principio, toda la distribución de muestreo de W se puede obtener dada la distribución probabilística de Y_i y la función h. Suele ser más fácil enfocarse en algunas características de la distribución de W cuando se evalúa como un estimador de θ. La primera propiedad importante de un estimador implica su valor esperado.

Estimador insesgado. Un estimador, W de θ, es un **estimador insesgado** si

$$\mathrm{E}(W) = \theta, \tag{C.3}$$

para todos los valores posibles de θ.

Si un estimador es insesgado, entonces su distribución de probabilidad tiene un valor esperado igual al parámetro que se supone estimará. La insesgadez *no* significa que la estimación que se obtuvo una muestra particular sea igual a θ, o incluso muy cercana a θ. Por el contrario, si se pudieran sacar *indefinidamente* muestras aleatorias en Y de la población, calcular una estimación cada vez y después promediarlas en todas las muestras aleatorias, se obtendría θ. Este experimento mental es abstracto, pues en la mayoría de las aplicaciones, sólo se tiene una muestra aleatoria con la cual trabajar.

Para un estimador que no es insesgado, se define su sesgo de la siguiente manera.

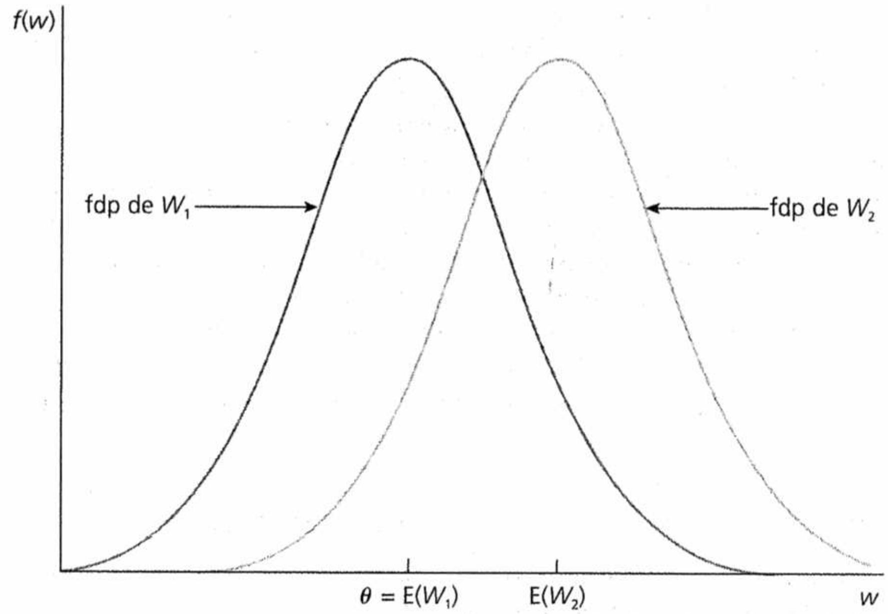
Sesgo de un estimador. Si W es un **estimador sesgado** de θ, su **sesgo** se define como

$$\mathrm{Sesgo}(W) \equiv \mathrm{E}(W) - \theta. \tag{C.4}$$

La figura C.1 muestra dos estimadores; el primero es insesgado y el segundo tiene un sesgo positivo.

FIGURA C.1

Un estimador insesgado, W_1, y un estimador con sesgo positivo, W_2.



La insesgadez de un estimador y la dimensión de cualquier sesgo posible dependen de la distribución de Y y de la función h. La distribución de Y suele estar fuera de control (aunque se suele elegir un *modelo* para esta distribución): puede estar determinado por la naturaleza o fuerzas sociales. Pero se tiene la opción de la regla h, y si se quiere un estimador, entónces se debe elegir h.

Por lo general, algunos estimadores pueden mostrar insesgadez. Por ahora se mostrará que el promedio muestral $\bar{Y}$ es un estimador insesgado de la media poblacional μ, sin importar la distribución de la población subyacente. Se utilizan las propiedades de los valores esperados (E.1 y E.2) que se cubrieron en la sección B.3:

$$\mathrm{E}(\bar{Y}) = \mathrm{E}\left((1/n)\sum_{i=1}^{n} Y_i\right) = (1/n)\mathrm{E}\left(\sum_{i=1}^{n} Y_i\right) = (1/n)\left(\sum_{i=1}^{n} \mathrm{E}(Y_i)\right)$$

$$= (1/n)\left(\sum_{i=1}^{n} \mu\right) = (1/n)(n\mu) = \mu.$$

Para probar la hipótesis, se necesita estimar la varianza σ^2 de una población con media μ. Si $\{Y_1, \dots, Y_n\}$ denota la muestra aleatoria de la población con $\mathrm{E}(Y) = \mu$ y $\mathrm{Var}(Y) = \sigma^2$, defina el estimador como

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(Y_i - \bar{Y})^2, \qquad \text{C.5}$$

lo cual suele recibir el nombre de **varianza muestral**. Se puede mostrar que S^2 es insesgada para σ^2: $E(S^2) = \sigma^2$. La división entre $n - 1$, en lugar de n, se debe al hecho de que la media μ es estimada, más que conocida. Si se conociera μ un estimador insesgado de σ^2 sería $n^{-1}\sum_{i=1}^{n}(Y_i - \mu)^2$, pero μ rara vez se conoce en la práctica.

Aunque la insesgadez tiene un cierto atractivo como propiedad para un estimador, de hecho, su antónimo, "sesgado", tiene fuertes connotaciones negativas, no está exento de problemas. Una debilidad de la insesgadez es que algunos estimadores razonables, e incluso algunos muy buenos, no son insesgados. Un poco más adelante se verá un ejemplo.

Otra debilidad importante de la insesgadez es que existen estimadores insesgados que en realidad son muy pobres. Considere estimar la media μ de una población. En lugar de utilizar el promedio muestral $\bar{Y}$ para estimar μ, suponga que, después de recabar una muestra del tamaño n, se descartan todas las observaciones salvo la primera. Es decir, el estimador de μ es simplemente $W \equiv Y_1$. Este estimador es insesgado debido a que $E(Y_1) = \mu$. Con algo de suerte, se piensa que ignorar todo excepto la primera observación, no es un método prudente para estimar: desecha la mayoría de la información en la muestra. Por ejemplo, con $n = 100$, se obtienen 100 resultados de la variable aleatoria Y, pero entonces se emplea sólo la primera de ellas para estimar $E(Y)$.

La varianza de muestreo de los estimadores

El ejemplo al final de la subsección previa muestra que se necesitan criterios adicionales para evaluar los estimadores. La insesgadez o insesgamiento sólo asegura que la distribución muestral de un estimador tenga un valor medio igual al parámetro que se supone va a estimar. Esto está bien, pero también es necesario saber qué tan dispersa es la distribución de un estimador. Un estimador puede ser igual a θ, en promedio, pero también puede estar muy lejos con una probabilidad muy alta. En la figura C.2, W_1 y W_2 son estimadores insesgados de θ. Pero la distribución de W_1 está más centrada en torno a θ: la probabilidad de que W_1 sea mayor que cualquier distancia dada a partir de θ es menor que la probabilidad de que W_2 sea mayor que la misma distancia a partir de θ. Utilizar W_1 como estimador significa que es menos probable que se obtenga una muestra aleatoria que produzca una estimación muy alejada de θ.

Para resumir la situación que se muestra en la figura C2, se utiliza la varianza (o desviación estándar) de un estimador. Recuerde que esto da una sola medida de la dispersión en la distribución. La varianza de un estimador suele recibir el nombre de **varianza de muestreo**, pues es la varianza asociada con una distribución muestral. Recuerde que la varianza de muestreo no es una variable aleatoria; sino una constante, aunque podría ser desconocida.

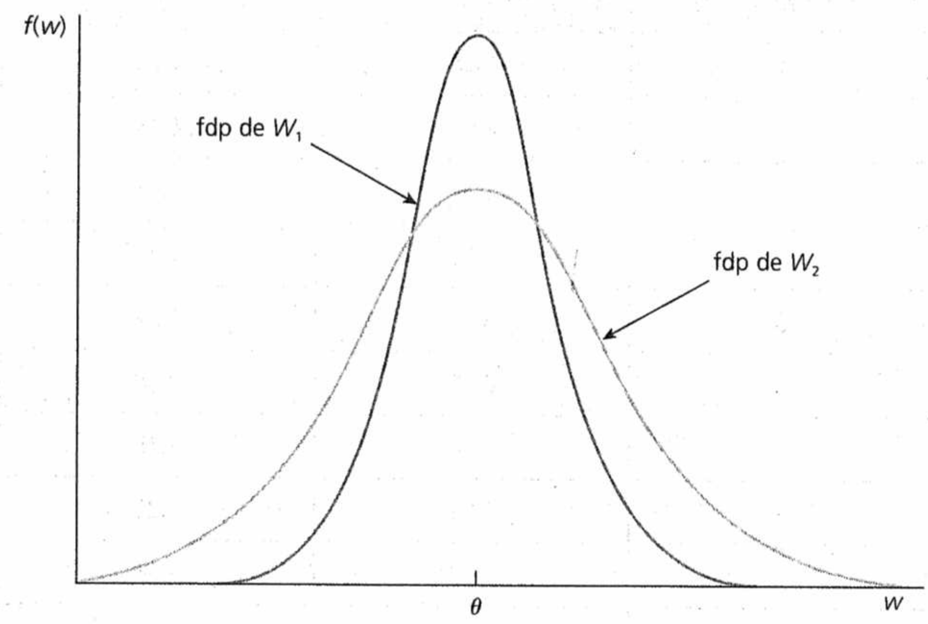
Se puede obtener ahora la varianza del promedio muestral para estimar la media μ de una población:

$$\begin{aligned}\mathrm{Var}(\bar{Y}) &= \mathrm{Var}\left((1/n)\sum_{i=1}^{n} Y_i\right) = (1/n^2)\mathrm{Var}\left(\sum_{i=1}^{n} Y_i\right) = (1/n^2)\left(\sum_{i=1}^{n} \mathrm{Var}(Y_i)\right) \\ &= (1/n^2)\left(\sum_{i=1}^{n} \sigma^2\right) = (1/n^2)(n\sigma^2) = \sigma^2/n.\end{aligned} \tag{C.6}$$

Observe cómo se emplean las propiedades de la varianza de las secciones B.3 y B.4 (VAR.2 y VAR.4), así como la independencia de Y_i. En suma: si $\{Y_i: i = 1, 2, \ldots, n\}$ es una muestra aleatoria de una población con una media μ y una varianza σ^2, entonces $\bar{Y}$ tiene la misma media que la población, pero su varianza muestral es igual a la varianza poblacional, σ^2, dividida entre el tamaño de la muestra.

FIGURA C.2

Distribuciones de dos estimadores insesgados de θ.



Una implicación importante de $\text{Var}(\bar{Y}) = \sigma^2/n$ es que puede acercarse mucho a cero al incrementar el tamaño muestral n. Esta es la característica clave de un estimador razonable, que se analizará en la sección C.3.

Como lo sugiere la figura C.2, entre los estimadores insesgados se prefiere el estimador con la varianza menor. Esto permite calcular ciertos estimadores de consideración. Para una muestra aleatoria de una población con media μ y varianza σ^2, se sabe que $\bar{Y}$ es insesgada y $\text{Var}(\bar{Y}) = \sigma^2/n$. ¿Qué hay del estimador Y_1, que es tan sólo la primera observación extraída? Debido a que Y_1 es una extracción aleatoria de la población, $\text{Var}(Y_1) = \sigma^2$. Por tanto, las diferencias entre $\text{Var}(Y_1)$ y $\text{Var}(\bar{Y})$ pueden ser grandes incluso para tamaños muestrales pequeños. Si $n = 10$, entonces $\text{Var}(Y_1)$ es 10 veces tan grande como $\text{Var}(\bar{Y}) = \sigma^2/10$. Esto da una forma formal de excluir Y_1 como estimador de μ.

Para enfatizar este punto, la tabla C.1 contiene el resultado de un pequeño estudio de simulación. Mediante el paquete estadístico Stata®, 20 muestras aleatorias de tamaño 10 se generaron de una distribución normal, con $\mu = 2$ y $\sigma^2 = 1$; aquí lo que interesa es estimar μ. Para cada una de las 20 muestras aleatorias se calculan dos estimaciones, y_1 y $\bar{y}$; estos valores están listados en la tabla C.1. Como se puede ver en la tabla, los valores de y_1 están mucho más dispersos que los de los rangos de $\bar{y}$: y_1 de -0.64 a 4.27, mientras que los rangos $\bar{y}$ sólo de 1.16 a 2.58. Además, en 16 de cada 20 casos, $\bar{y}$ es más cercana que y_1 a $\mu = 2$. El promedio de y_1 a través de las simulaciones es aproximadamente 1.89, mientras que el de $\bar{y}$ es 1.96. El hecho de que estos promedios sean cercanos a 2 ilustra el insesgamiento de ambos estimadores (y se podría hacer que estos promedios se acercaran más a 2 al hacer más de 20 réplicas). Pero comparar sólo los resultados promedio a través de las extracciones aleatorias oculta el hecho de que el promedio muestral $\bar{Y}$ es muy superior a Y_1 como estimador de μ.

TABLA C.1

Simulación de estimadores para una distribución Normal(μ,1) con μ = 2

Réplica	y_1	$\bar{y}$
1	−0.64	1.98
2	1.06	1.43
3	4.27	1.65
4	1.03	1.88
5	3.16	2.34
6	2.77	2.58
7	1.68	1.58
8	2.98	2.23
9	2.25	1.96
10	2.04	2.11
11	0.95	2.15
12	1.36	1.93
13	2.62	2.02
14	2.97	2.10
15	1.93	2.18
16	1.14	2.10
17	2.08	1.94
18	1.52	2.21
19	1.33	1.16
20	1.21	1.75

Eficiencia

Si se comparan las varianzas de $\bar{Y}$ y Y_1 de la sección anterior, se obtiene un ejemplo de un método general para comparar estimadores insesgados.

Eficiencia relativa. Si W_1 y W_2 son dos estimadores insesgados de θ, W_1 es eficiente en relación con W_2 cuando $\text{Var}(W_1) \leq \text{Var}(W_2)$ para toda θ, con desigualdad estricta al menos para un valor de θ.

Antes mostramos que, para estimar la media poblacional μ, $\text{Var}(\bar{Y}) < \text{Var}(Y_1)$ para cualquier valor de σ^2 siempre que $n > 1$. Por tanto, $\bar{Y}$ es eficiente en relación con Y_1 para estimar μ. No siempre se puede elegir entre estimadores insesgados con base en el criterio de menor varianza: dados dos estimadores insesgados de θ, uno puede obtener una varianza menor de algunos valores de θ, mientras que el otro puede tener una varianza menor para otros valores de θ.

Si se restringe la atención a ciertas clases de estimadores, se puede mostrar que el promedio muestral tiene la varianza más pequeña. El problema C.2 pide mostrar que $\bar{Y}$ tiene la menor varianza entre todos los estimadores insesgados que son también funciones lineales de Y_1, Y_2, …, Y_n. Los supuestos consisten en que Y_i tiene una media y varianza común y que no están correlacionados a pares.

Si no se restringe la atención a estimadores insesgados, entonces no tiene sentido comparar las varianzas. Por ejemplo, cuando se estima la media poblacional μ, se puede usar un estimador trivial que es igual a cero, sin importar la muestra que se extraiga. Como es natural, la varianza de este estimador es cero (dado que es el mismo valor para toda muestra aleatoria). Pero el sesgo del estimador es $-\mu$, así que es un estimador muy pobre cuando $|\mu|$ es grande.

Una forma de comparar los estimadores que no necesariamente son insesgados, es calcular el **error cuadrático medio (ECM)** de los estimadores. Si W es un estimador de θ, entonces el ECM de W se define como $\text{ECM}(W) = \text{E}[(W - \theta)^2]$. El ECM mide, en promedio, qué tan alejado está el estimador de θ. Se puede mostrar que $\text{ECM}(W) = \text{Var}(W) + [\text{Bias}(W)]^2$, así que $\text{ECM}(W)$ depende de la varianza y el sesgo (si lo hubiera). Esto permite comparar dos estimadores cuando uno o ambos son sesgados.

C.3 Propiedades asintóticas o de muestra grande de los estimadores

En la sección C.2 se encontró el estimador Y_1 para la media poblacional μ, y se vio que aunque es insesgado, es un estimador pobre debido a que su varianza puede ser mucho más grande que la media muestral. Una característica notable de Y_1 es que tiene la misma varianza para cualquier tamaño de muestra. Parece razonable exigir que el tamaño de la muestra aumente para que cualquier procedimiento de estimación mejore. Para estimar una media poblacional μ, $\bar{Y}$ mejora en el sentido de que la varianza disminuye a medida que n crece; Y_1 no mejora en este sentido.

Se pueden descartar ciertos estimadores inútiles al estudiar las propiedades *asintóticas* o de *muestra grande* de los estimadores. Además, se puede decir algo positivo acerca de los estimadores que no son insesgados y cuyas varianzas no son fáciles de encontrar.

El análisis asintótico implica aproximar las características de la distribución muestral de un estimador. Estas aproximaciones dependen del tamaño de la muestra. Por desgracia, necesariamente hay un límite en lo que se puede decir acerca de qué tan "grande" necesita ser un tamaño muestral para que el análisis asintótico sea apropiado; esto depende de la distribución de la población subyacente. Pero se sabe que las aproximaciones de muestras grandes funcionan bien para tamaños muestrales tan pequeños como $n = 20$.

Consistencia

La primera propiedad asintótica de los estimadores concierne a qué tan lejos es probable que esté el estimador del parámetro que se supone estimará, a medida que el tamaño de la muestra aumenta indefinidamente.

Consistencia. Sea W_n un estimador de θ basado en una muestra $Y_1, Y_2, \ldots, Y_n$ del tamaño n. Entonces, W_n es un **estimador consistente** de θ si para toda $\varepsilon > 0$,

$$P(|W_n - \theta| > \varepsilon) \to 0 \text{ cuando } n \to \infty. \tag{C.7}$$

Si W_n no es consistente para θ, entonces se dice que es **inconsistente**.

Cuando W_n es consistente, también se dice que θ es el **límite en probabilidad** de W_n, escrito como $\text{plim}(W_n) = \theta$.

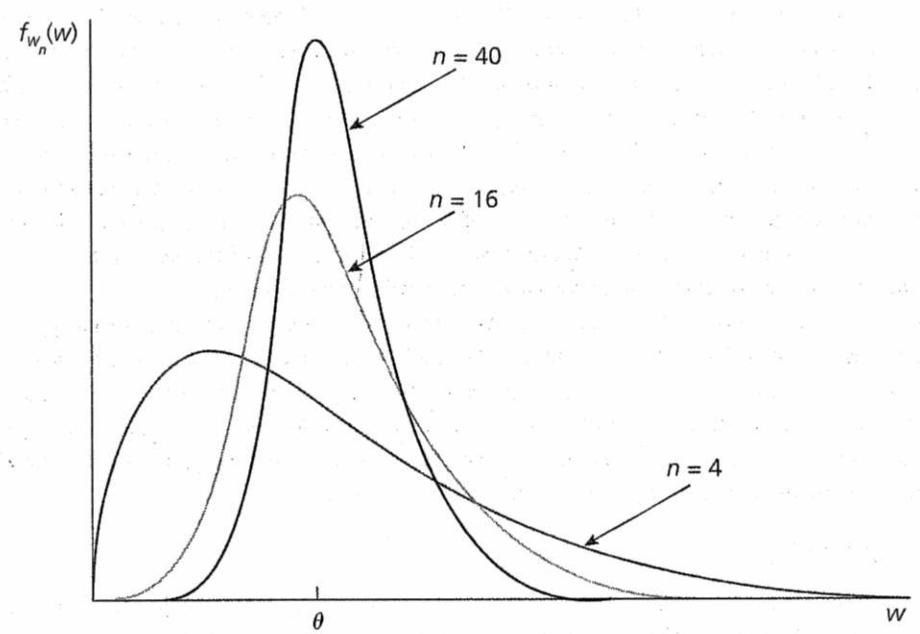
Al igual que el insesgamiento, que es una característica de un estimador de un tamaño de la muestra determinado, la consistencia implica el comportamiento de la distribución muestral de un estimador a medida que el tamaño muestral n aumenta. Para enfatizar esto, en esta definición, se ha indexado el estimador por el tamaño de la muestra y se continuará con esta convención a lo largo de esta sección.

La ecuación (C.7) parece técnica y puede ser muy difícil de establecer con base en principios fundamentales de probabilidad. Por el contrario, interpretar (C.7) es fácil. Significa que la distribución de W_n se concentra cada vez más en torno de θ, lo cual a grandes rasgos, significa que para tamaños de la muestra mayores, es cada vez menos probable que W_n se aleje mucho de θ. Esta tendencia se ilustra en la figura C.3.

Si un estimador no es consistente, entonces no es útil saber de θ, incluso con una cantidad ilimitada de datos. Por esta razón, la consistencia es un requisito mínimo de un estimador utilizado en estadística o econometría. Se encontrarán estimadores consistentes bajo ciertos supuestos e inconsistentes cuando tales supuestos fallen. Cuando los estimadores son inconsistentes, por

FIGURA C.3

Distribuciones de muestreo de un estimador consistente para tres tamaños de muestras.



lo general, se encuentran sus límites de probabilidad y será importante saber qué tan lejos se encuentran éstos de θ.

Como se observó antes, los estimadores insesgados no son necesariamente consistentes, pero aquellos con varianzas que se reducen a cero a medida que el tamaño muestral aumenta *son* consistentes. Esto se puede expresar formalmente: si W_n es un estimador insesgado de θ y $\text{Var}(W_n) \rightarrow 0$ cuando $n \rightarrow \infty$, entonces $\text{plim}(W_n) = \theta$. Los estimadores insesgados que utilizan la muestra de datos completa, por lo general, tienen una varianza que se reduce a cero a medida que el tamaño muestral aumenta y, por tanto, son consistentes.

Un buen ejemplo de un estimador consistente es el promedio de una muestra aleatoria extraída de la población con media μ y varianza σ^2. Ya se ha mostrado que el promedio muestral es insesgado para μ. En la ecuación (C.6) se derivó $\text{Var}(\bar{Y}_n) = \sigma^2/n$ para cualquier tamaño de muestra n. Por tanto, $\text{Var}(\bar{Y}_n) \rightarrow 0$ cuando $n \rightarrow \infty$, así, $\bar{Y}_n$ es un estimador consistente de μ (además de ser insesgado).

La conclusión de que $\bar{Y}_n$ es consistente para μ es válida incluso si $\text{Var}(\bar{Y}_n)$ no existiera. Este resultado clásico se conoce como la **ley de los grandes números (LGN)**.

Ley de los grandes números. Sean $Y_1, Y_2, \ldots, Y_n$ variables aleatorias independientes, idénticamente distribuidas con media μ. Entonces,

$$\text{plim}(\bar{Y}_n) = \mu. \qquad \boxed{\text{C.8}}$$

La ley de los grandes números significa que, si interesa estimar el promedio poblacional μ, es posible aproximarse arbitrariamente a μ si se elige una muestra suficientemente grande. Este resultado fundamental se puede combinar con las propiedades básicas de los plim para mostrar que estimadores muy complejos son consistentes.

Propiedad PLIM.1: Sea θ un parámetro y defina un nuevo parámetro, $\gamma = g(\theta)$, para alguna función *continua* $g(\theta)$. Suponga que $\text{plim}(W_n) = \theta$. Defina un estimador de γ por $G_n = g(W_n)$. Entonces,

$$\text{plim}(G_n) = \gamma. \qquad \boxed{\text{C.9}}$$

Esto suele expresarse como

$$\text{plim}\, g(W_n) = g(\text{plim}\, W_n) \qquad \boxed{\text{C.10}}$$

para una función continua $g(\theta)$.

El supuesto de que $g(\theta)$ es continua, es un requisito técnico que suele describirse de forma no técnica como "una función que puede graficarse sin levantar el lápiz del papel". Debido a que todas las funciones que se encuentran en este libro son continuas, no se ofrece una definición formal de una función continua. Ejemplos de funciones continuas son $g(\theta) = a + b\theta$ para constantes a y b, $g(\theta) = \theta^2$, $g(\theta) = 1/\theta$, $g(\theta) = \sqrt{\theta}$, $g(\theta) = \exp(\theta)$, y muchas variantes de éstas. No será necesario mencionar nuevamente el supuesto de continuidad.

Como ejemplo importante de un estimador consistente pero sesgado, considere la desviación estándar, σ, de una población con media μ y varianza σ^2. Se ha afirmado que la varianza muestral $S_n^2 = (n-1)^{-1}\sum_{i=1}^{n}(Y_i - \bar{Y}_n)^2$ es insesgada para σ^2. Mediante la ley de los grandes números y un poco de álgebra, S_n^2 se puede mostrar que es consistente para σ^2. El estimador natural de

$\sigma = \sqrt{\sigma^2}$ es $S_n = \sqrt{S_n^2}$ (donde la raíz cuadrada es siempre la raíz cuadrada positiva). S_n, que recibe el nombre de **desviación estándar muestral**, *no* es un estimador insesgado debido a que el valor esperado de la raíz cuadrada *no* es la raíz cuadrada del valor esperado (vea la sección B.3). No obstante, por PLIM.1, plim $S_n = \sqrt{\text{plim } S_n^2} = \sqrt{\sigma^2} = \sigma$, así que S_n es un estimador consistente de σ.

He aquí otras propiedades útiles del límite en probabilidad:

Propiedad PLIM.2: Si $\text{plim}(T_n) = \alpha$ y $\text{plim}(U_n) = \beta$, entonces

i) $\text{plim}(T_n + U_n) = \alpha + \beta$;

ii) $\text{plim}(T_n U_n) = \alpha\beta$;

iii) $\text{plim}(T_n / U_n) = \alpha/\beta$, siempre que $\beta \neq 0$.

Estas tres propiedades sobre los límites en probabilidad permiten combinar estimadores consistentes en varias formas para obtener otros estimadores consistentes. Por ejemplo, sea $\{Y_1, \ldots, Y_n\}$ una muestra aleatoria de tamaño n sobre los ingresos anuales de la población de trabajadores con una educación de bachillerato y denótese la media poblacional por μ_Y. Sea $\{Z_1, \ldots, Z_n\}$ una muestra aleatoria del ingreso anual de la población de trabajadores con una educación universitaria y denote la media poblacional con μ_Z. Se quiere estimar la diferencia porcentual en el ingreso anual entre los dos grupos, la cual es $\gamma = 100 \cdot (\mu_Z - \mu_Y)/\mu_Y$. (Este es el porcentaje mediante el cual el ingreso promedio de los graduados universitarios difiere del ingreso promedio para los graduados de bachillerato.) Debido a que $\bar{Y}_n$ es consistente para μ_Y y $\bar{Z}_n$ es consistente para μ_Z, se desprende de PLIM.1 y el inciso iii) de PLIM.2 que

$$G_n \equiv 100 \cdot (\bar{Z}_n - \bar{Y}_n)/\bar{Y}_n$$

es un estimador consistente de γ. G_n es sólo la diferencia porcentual entre $\bar{Z}_n$ y $\bar{Y}_n$ en la muestra, así que es un estimador natural. G_n no es un estimador insesgado de γ, pero sigue siendo un buen estimador salvo, quizá, cuando n es pequeña.

Normalidad asintótica

La consistencia es una propiedad de los estimadores puntuales. Aunque ésta expresa que la distribución del estimador se está colapsando en torno al parámetro a medida que el tamaño de la muestra aumenta, no expresa esencialmente nada acerca de la *forma* de esa distribución para un tamaño muestral dado. Para construir estimadores de intervalo y pruebas de hipótesis, se necesita una forma de aproximar la distribución de los estimadores. La mayoría de los estimadores econométricos tiene distribuciones que se aproximan bien mediante una distribución normal para muestras grandes, lo que motiva la siguiente definición.

Normalidad asintótica. Sea $\{Z_n: n = 1, 2, \ldots\}$ una secuencia de variables aleatorias, tal que para todos los números z,

$$\text{P}(Z_n \leq z) \rightarrow \Phi(z) \text{ as } n \rightarrow \infty, \qquad \boxed{\textbf{C.11}}$$

donde $\Phi(z)$ es la función de distribución acumulada normal estándar. Entonces, se dice que, Z_n tiene una *distribución normal estándar asintótica*. En este caso suele escribirse $Z_n \overset{a}{\sim} \text{Normal}(0,1)$. (La "$a$" sobre la tilde significa "asintóticamente" o "aproximadamente".)

La propiedad (C.11) significa que la función de Z_n se acerca cada vez más a la fda de la distribución normal estándar a medida que el tamaño de la muestra n aumenta. Cuando la **normalidad**

asintótica se mantiene para n grandes se tiene la aproximación $P(Z_n \leq z) \approx \Phi(z)$. Por tanto, las probabilidades concernientes a Z_n se pueden aproximar mediante probabilidades normales estándar.

El **teorema del límite central (TLC)** es uno de los resultados más poderosos en probabilidad y estadística. Indica que el promedio de una muestra aleatoria para *cualquier* población (con varianza finita), cuando se estandariza, tiene una distribución normal estándar asintótica.

Teorema del límite central. Sea $\{Y_1, Y_2, \ldots, Y_n\}$ una muestra aleatoria con una media μ y una varianza σ^2. Entonces,

$$Z_n = \frac{\bar{Y}_n - \mu}{\sigma/\sqrt{n}} \qquad \boxed{\textbf{C.12}}$$

tiene una distribución normal estándar asintótica.

La variable Z_n en (C.12) es la versión estandarizada de $\bar{Y}_n$: se ha sustraído $E(\bar{Y}_n) = \mu$ y dividido entre $sd(\bar{Y}_n) = \sigma/\sqrt{n}$. Por tanto, sin importar la distribución poblacional de Y, Z_n tiene media cero y varianza uno, lo cual coincide con la media y varianza de la distribución normal estándar. Es de notar que toda la distribución de Z_n se acerca arbitrariamente a la distribución normal estándar a medida que n aumenta.

Se puede escribir la variable estandarizada en la ecuación (C.12) como $\sqrt{n}(\bar{Y}_n - \mu)/\sigma$, lo cual muestra que se debe multiplicar la diferencia entre la media muestral y la media poblacional por la raíz cuadrada del tamaño de la muestra para obtener una distribución limitante útil. Sin la multiplicación por $\sqrt{n}$, se tendría sólo $(\bar{Y}_n - \mu)/\sigma$, que converge en probabilidad a cero. En otras palabras, la distribución de $(\bar{Y}_n - \mu)/\sigma$ simplemente se colapsa a un solo punto conforme $n \to \infty$, lo cual, se sabe, no puede ser una buena aproximación a la distribución de $(\bar{Y}_n - \mu)/\sigma$ para tamaños muestrales razonables. Multiplicar por $\sqrt{n}$ asegura que la varianza de Z_n permanece constante. Prácticamente, suele tratarse a $\bar{Y}_n$ como distribuida normal y aproximadamente con media μ y varianza σ^2/n, y esto provee el procedimiento estadístico correcto, debido a que lleva a la variable estandarizada en la ecuación (C.12).

La mayoría de los estimadores encontrados en estadística y econometría se puede escribir como funciones de promedios muestrales, en cuyo caso se puede aplicar la ley de los grandes números y el teorema del límite central. Cuando dos estimadores consistentes tiene distribuciones normales asintóticas, se elige el estimador con la varianza asintótica menor.

Además, el promedio muestral estandarizado en (C.12), muchos otros estadísticos que dependen de promedios muestrales resultan ser asintóticamente normales. Uno importante es el que se obtiene al reemplazar σ con su estimador consistente S_n en la ecuación (C1.12):

$$\frac{\bar{Y}_n - \mu}{S_n/\sqrt{n}} \qquad \boxed{\textbf{C.13}}$$

también tiene una distribución normal estándar aproximada para n grande. Las distribuciones exactas (muestra finita) de (C.12) y (C.13) definitivamente no son iguales, pero la diferencia suele ser tan pequeña como para ignorarse cuando n es grande.

A lo largo de esta sección, cada estimador tiene un subíndice n para enfatizar la naturaleza del análisis muestral grande o asintótico. Continuar esta convención desorganizaría la notación sin ofrecer información adicional, una vez que se comprendieran los principios básicos del análisis asintótico. Por tanto, se descartará el subíndice n, con la esperanza de que se recuerde que los estimadores dependen del tamaño muestral, y las propiedades como consistencia y normalidad asintótica se refieren al crecimiento sin límite del tamaño de la muestra.

C.4 Métodos generales para estimar parámetros

Hasta este punto se ha utilizado el promedio muestral para ilustrar las propiedades finitas y de muestra grande de los estimadores. Es natural preguntar: ¿existen métodos generales para estimar que produzcan estimadores con buenas propiedades, como insesgadez, consistencia y eficiencia?

La respuesta es sí. Un análisis detallado de los diversos métodos que existen para estimar está más allá del alcance de este libro; aquí sólo se ofrecerá un análisis informal. Un análisis detallado se puede encontrar en Larsen y Marx (1986, capítulo 5).

El método de momentos

Dado un parámetro θ que aparece en una distribución poblacional, suele haber diversas formas de obtener estimadores insesgados y consistentes de θ. No sería práctico, intentar todas las diferentes posibilidades y compararlas con base en los criterios de las secciones C.2 y C.3. Por fortuna, se ha mostrado que algunos métodos tienen buenas propiedades generales y, en la mayor parte, la lógica en la que se basan es atractiva, desde un punto de vista intuitivo.

En secciones anteriores, se ha estudiado el promedio muestral como un estimador insesgado del promedio poblacional y la varianza muestral como un estimador insesgado de la varianza poblacional. Estos estimadores son ejemplos de los estimadores del **método de momentos**. En general, la estimación del método de momentos es la siguiente. El parámetro θ está relacionado con algún valor esperado en la distribución de Y, que suele ser $\mathrm{E}(Y)$ o $\mathrm{E}(Y^2)$ (aunque en ocasiones se utilizan opciones más exóticas). Suponga, por ejemplo, que el parámetro de interés, θ, está relacionado con la media poblacional como $\theta = g(\mu)$ para alguna función g. Debido a que el promedio muestral $\bar{Y}$ es insesgado y es un estimador consistente de μ, es natural remplazar μ con $\bar{Y}$, lo cual genera un estimador $g(\bar{Y})$ de θ. El estimador $g(\bar{Y})$ es consistente para θ y si $g(\mu)$ es una función lineal de μ, entonces $g(\bar{Y})$ es insesgado también. Lo que se ha hecho es remplazar el momento poblacional, μ, con su contraparte muestral, $\bar{Y}$. De ahí el nombre de "método de momentos".

Se abarcan dos estimadores adicionales del método de momentos que serán útiles para la discusión del análisis de momentos. Recuerde que la covarianza entre dos variables aleatorias X y Y se define como $\sigma_{XY} = \mathrm{E}[(X - \mu_X)(Y - \mu_Y)]$. El método de momentos sugiere estimar σ_{XY} por $n^{-1}\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})$. Este es un estimador consistente de σ_{XY}, pero resulta ser sesgado esencialmente por la misma razón de que la varianza muestral es sesgada si n, y no $n - 1$, se usa como el divisor. La **covarianza muestral** se define como

$$S_{XY} = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y}). \qquad \boxed{\textbf{C.14}}$$

Se puede mostrar que este es un estimador insesgado de σ_{XY}. (Remplazar n con $n - 1$ no produce diferencia a medida que el tamaño muestral aumenta de forma indefinida, así que este estimador sigue siendo consistente).

Como se analizó en la sección B.4, la covarianza entre dos variables suele ser difícil de interpretar. Por lo general, lo que interesa más es la correlación. Debido a que la correlación poblacional es $\rho_{XY} = \sigma_{XY}/(\sigma_X\sigma_Y)$, el método de momentos sugiere que estimar ρ_{XY} como

$$R_{XY} = \frac{S_{XY}}{S_X S_Y} = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{\left(\sum_{i=1}^{n}(X_i - \bar{X})^2\right)^{1/2}\left(\sum_{i=1}^{n}(Y_i - \bar{Y})^2\right)^{1/2}}, \qquad \boxed{\textbf{C.15}}$$

lo cual recibe el nombre de **coeficiente de correlación muestral** (o *correlación muestral*). Observe que se ha cancelado la división mediante $n - 1$ en la covarianza muestral y las desviaciones estándar muestrales. De hecho, se puede dividir cada una entre n, y se llegaría finalmente a la misma fórmula.

Se puede mostrar que el coeficiente de correlación muestral siempre está en el intervalo $[-1,1]$, como debe ser. Debido a que S_{XY}, S_X y S_Y son consistentes para el parámetro poblacional correspondiente, R_{XY} es un estimador consistente de la correlación poblacional, ρ_{XY}. No obstante, R_{XY} es un estimador sesgado por dos razones. Primero, S_X y S_Y son estimadores sesgados de σ_X y σ_Y, respectivamente. Segundo, R_{XY} es una razón de estimadores, así que no puede ser insesgado, aunque S_X y S_Y lo fueran. Para los fines de este libro, esto no es importante, a pesar del hecho de que un resultado clásico en la estadística matemática es que no existe un estimador insesgado de ρ_{XY}.

Máxima verosimilitud

Otro método general para estimar es el método de *máxima verosimilitud*, un tema que se cubre en muchos cursos introductorios de estadística. Un breve resumen sobre el caso más simple bastará aquí. Sea $\{Y_1, Y_2, \ldots, Y_n\}$ una muestra aleatoria de la distribución poblacional $f(y;\theta)$. Debido al supuesto de muestreo aleatorio, la distribución conjunta de $\{Y_1, Y_2, \ldots, Y_n\}$ es simplemente el producto de las densidades: $f(y_1;\theta)f(y_2;\theta) \cdots f(y_n;\theta)$. En el caso discreto, éste es $P(Y_1 = y_1, Y_2 = y_2, \ldots, Y_n = y_n)$. Ahora, defínase la *función de verosimilitud* como

$$L(\theta;Y_1, \ldots, Y_n) = f(Y_1;\theta)f(Y_2;\theta) \cdots f(Y_n;\theta),$$

que es una variable aleatoria debido a que depende del resultado de la muestra aleatoria $\{Y_1, Y_2, \ldots, Y_n\}$. El **estimador de máxima verosimilitud** de θ, llámese W, es el valor de θ que maximiza la función de verosimilitud. (Ésta es la razón de que se escriba L en función de θ, seguida por la muestra aleatoria). Por supuesto, este valor depende de la muestra aleatoria. El principio de máxima verosimilitud dice que, de todos los valores posibles de θ, se debe elegir el valor que haga que la verosimilitud de los datos observados sea mayor. Intuitivamente, éste es un método razonable para estimar θ.

Por lo general, es más conveniente trabajar con la *función log-probabilidad*, que se obtiene mediante el logaritmo natural de la función de verosimilitud:

$$\log [L(\theta; Y_1, \ldots, Y_n)] = \sum_{i=1}^{n} \log [f(Y_i; \theta)], \qquad \boxed{\textbf{C.16}}$$

donde se utiliza el hecho de que el logaritmo del producto es la suma de los logaritmos. Dado que (C.16) es la suma de variables aleatorias idénticamente distribuidas e independientes, analizar los estimadores que provienen de (C.16) es relativamente fácil.

La estimación de máxima verosimilitud (EMV) suele ser consistente y, en ocasiones, insesgada. Pero existen muchos otros estimadores. El atractivo tan difundido de EMV se debe a que, en general es el estimador más asintóticamente eficiente cuando el modelo poblacional $f(y;\theta)$ se especifica correctamente. Además, el EMV algunas veces es el **estimador insesgado de varianza mínima**, es decir, tiene la varianza menor entre todos los estimadores insesgados de θ. [Vea Larsen y Marx (1986, capítulo 5) para verificar estas afirmaciones.]

En el capítulo 17 se necesitará una máxima verosimilitud para estimar los parámetros de más modelos econométricos avanzados. En econometría casi siempre se tiene interés en la distribución de Y condicional en un conjunto de variables explicativas, por ejemplo, $X_1, X_2, \ldots, X_k$. Entonces, se remplaza la densidad en (C.16) con $f(Y_i|X_{i1}, \ldots, X_{ik}; \theta_1, \ldots, \theta_p)$, donde a esta densidad se le permite depender de p parámetros, $\theta_1, \ldots, \theta_p$. Por fortuna, para la aplicación exitosa de los métodos de máxima verosimilitud, no es necesaria una mayor profundización en las cuestiones computacionales o teorías estadísticas de muestra grande. Wooldridge (2002, capítulo 13) cubre la teoría de la estimación de máxima verosimilitud.

Mínimos cuadrados

Un tercer tipo de estimador, y que desempeña la mayor de las funciones en el libro, es el llamado **estimador de mínimos cuadrados**. Ya se ha visto un ejemplo de mínimos cuadrados: la media muestral, $\bar{Y}$, es un estimador de mínimos cuadrados de la media poblacional, μ. Ya se sabe que $\bar{Y}$ es un estimador del método de momentos. ¿Qué lo convierte en un estimador de mínimos cuadrados? Se puede mostrar que el valor de m que compone la suma de las desviaciones cuadradas

$$\sum_{i=1}^{n}(Y_i - m)^2$$

tan pequeña como sea posible es $m = \bar{Y}$. Demostrar esto no es difícil, por lo que se omite el álgebra.

Para algunas distribuciones importantes, incluidas la normal y la de Bernoulli, el promedio muestral $\bar{Y}$ es también el estimador de máxima verosimilitud de la media poblacional μ. Por tanto, los principios de los mínimos cuadrados, método de momentos y máxima verosimilitud suelen generar el *mismo* estimador. En otros casos, los estimadores son similares, mas no idénticos.

C.5 Estimación de intervalos e intervalos de confianza

La naturaleza de la estimación de intervalos

Una estimación puntual obtenida de una muestra particular no proporciona, por sí misma, suficiente información para probar teorías económicas o para análisis de políticas. Una estimación puntual puede ser el mejor cálculo del valor poblacional, pero, por su naturaleza, no ofrece información sobre qué tanta "probabilidad" hay de que la estimación esté cercana al parámetro de la población. Por ejemplo, suponga que un investigador reporta, con base en una muestra aleatoria de trabajadores, que los subsidios a la capacitación laboral incrementan el salario por hora en 6.4%. ¿Cómo se sabrá si esto es semejante o no al efecto en la población de trabajadores que pudieran capacitarse? Debido a que se ignora el valor poblacional, no se puede saber qué tan próxima es una estimación para una muestra en particular. No obstante, se pueden hacer planteamientos que comprendan probabilidades, y aquí es donde entra la estimación de intervalo.

Ya se conoce una forma de evaluar la incertidumbre en un estimador: determinar su desviación estándar muestral. Revelar la desviación estándar del estimador, junto con la estimación puntual, ofrece cierta información sobre la precisión de la estimación en cuestión. No obstante, aunque se ignore el problema de la dependencia de la desviación estándar sobre los parámetros de población desconocidos, dar cuenta de la desviación estándar junto con la estimación puntual no ofrece ningún planteamiento directo sobre dónde es probable que el valor poblacional se ubique con relación a la estimación. Esta limitación se supera al construir el **intervalo de confianza**.

Se ilustra el concepto de intervalo de confianza con un ejemplo. Suponga que la población tiene una distribución Normal(μ,1) y sea $\{Y_1, \ldots, Y_n\}$ una muestra aleatoria de esta población. (Se supone que la varianza poblacional es conocida y que es igual a la unidad, para fines ilustrativos; más adelante se mostrará qué hacer en el caso más realista en que la varianza se ignora.)

El promedio muestral, $\bar{Y}$, tiene una distribución normal con una media μ y una varianza $1/n$: $\bar{Y} \sim \text{Normal}(\mu, 1/n)$. A partir de esto, se puede estandarizar $\bar{Y}$, y, debido a que la versión estándar de $\bar{Y}$ tiene una distribución normal estándar, se tiene

$$\mathrm{P}\left(-1.96 < \frac{\bar{Y} - \mu}{1/\sqrt{n}} < 1.96\right) = .95.$$

El evento entre paréntesis es idéntico al evento $\bar{Y} - 1.96/\sqrt{n} < \mu < \bar{Y} + 1.96/\sqrt{n}$, así que

$$\mathrm{P}(\bar{Y} - 1.96/\sqrt{n} < \mu < \bar{Y} + 1.96/\sqrt{n}) = .95. \qquad \textbf{C.17}$$

La ecuación (C.17) es interesante, debido a que indica que la probabilidad de que el intervalo aleatorio $[\bar{Y} - 1.96/\sqrt{n}, \bar{Y} + 1.96/\sqrt{n}]$ contenga la media poblacional μ es .95, o 95%. Esta información permite construir una *estimación de intervalo* de μ, lo cual se obtiene al insertar el resultado muestral del promedio, $\bar{y}$. Por tanto,

$$[\bar{y} - 1.96/\sqrt{n}, \bar{y} + 1.96/\sqrt{n}] \qquad \textbf{C.18}$$

es un ejemplo de una estimación de intervalo de μ. También recibe el nombre de un intervalo de confianza a 95%. Una notación abreviada para este intervalo es $\bar{y} \pm 1.96/\sqrt{n}$.

El intervalo de confianza en la ecuación (C.18) es fácil de calcular, una vez que se observan los datos muestrales $\{y_1, y_2, \ldots, y_n\}$; $\bar{y}$ es el único factor que depende de los datos. Por ejemplo, suponga que $n = 16$ y el promedio de las 16 observaciones es 7.3. Entonces, el intervalo de confianza a 95% para μ es $7.3 \pm 1.96/\sqrt{16} = 7.3 \pm .49$, lo cual se puede escribir en forma de intervalo como [6.81,7.79]. Por construcción, $\bar{y} = 7.3$ está en el centro de este intervalo.

A diferencia de este cálculo, el significado de intervalo de confianza es más difícil de comprender. Cuando se dice que la ecuación (C.18) está a un intervalo de confianza a 95% para μ, significa que el intervalo *aleatorio*

$$[\bar{Y} - 1.96/\sqrt{n}, \bar{Y} + 1.96/\sqrt{n}] \qquad \textbf{C.19}$$

contiene a μ con probabilidad de .95. En otras palabras, antes de que se extraiga la muestra aleatoria, existe una probabilidad de 95% de que (C.19) contenga a μ. La ecuación (C.19) es un ejemplo de un **estimador de intervalo**. Es un intervalo aleatorio, dado que los puntos finales cambian con diferentes muestras.

Un intervalo de confianza suele interpretarse de la siguiente manera: "La probabilidad de que μ esté en el intervalo (C.18) es de .95." Esto es incorrecto. Una vez que la muestra se observa y $\bar{y}$ se calcula, los límites del intervalo de confianza son simplemente números (6.81 y 7.79 en el ejemplo anterior). El parámetro poblacional, μ, aunque desconocido, es también algún otro número. Por tanto, μ está o no en el intervalo (C.18) (y nunca se sabrá con certidumbre cuál es el caso). La probabilidad no desempeña ninguna función una vez que el intervalo de confianza se ha calculado para los datos particulares a la mano. La interpretación probabilística proviene del hecho que para 95% de todas las muestras aleatorias, el intervalo de confianza construido contendrá a μ.

Para enfatizar el significado de un intervalo de confianza, la tabla C.2 contiene los cálculos para 20 muestras aleatorias (o réplicas) de la distribución Normal(2,1) con un tamaño muestral $n = 10$. Para cada una de las 20 muestras, se obtiene, $\bar{y}$ y (C.18) se calcula como $\bar{y} \pm 1.96/\sqrt{10} = \bar{y} \pm .62$ (cada uno de ellos redondeado a dos decimales). Como puede verse, el intervalo cambia con cada muestra aleatoria. Diecinueve de los 20 intervalos contienen al valor poblacional de μ. Sólo para la réplica número 19 μ no está en el intervalo de confianza. En otras palabras, 95% de las muestras genera un intervalo de confianza que contiene a μ. Esto no tendría que ser el caso con sólo 20 réplicas, pero así se ideó para esta simulación en particular.

TABLA C.2

Intervalos de confianza simulados de una distribución Normal(μ,1) con $\mu = 2$

Réplica	$\bar{y}$	Intervalo a 95%	¿Contiene a μ?
1	1.98	(1.36,2.60)	Sí
2	1.43	(0.81,2.05)	Sí
3	1.65	(1.03,2.27)	Sí
4	1.88	(1.26,2.50)	Sí
5	2.34	(1.72,2.96)	Sí
6	2.58	(1.96,3.20)	Sí
7	1.58	(.96,2.20)	Sí
8	2.23	(1.61,2.85)	Sí
9	1.96	(1.34,2.58)	Sí
10	2.11	(1.49,2.73)	Sí
11	2.15	(1.53,2.77)	Sí
12	1.93	(1.31,2.55)	Sí
13	2.02	(1.40,2.64)	Sí
14	2.10	(1.48,2.72)	Sí
15	2.18	(1.56,2.80)	Sí
16	2.10	(1.48,2.72)	Sí
17	1.94	(1.32,2.56)	Sí
18	2.21	(1.59,2.83)	Sí
19	1.16	(.54,1.78)	No
20	1.75	(1.13,2.37)	Sí

Intervalos de confianza para la media de una población normalmente distribuida

El intervalo de confianza derivado en la ecuación (C.18) ayuda a ilustrar cómo construir e interpretar intervalos de confianza. En la práctica, la ecuación (C.18) no es muy útil para la media de una población normal, pues se supone que el valor conocido de la varianza es uno. Es fácil extender (C.18) al caso en el que se sabe que la desviación estándar σ toma cualquier valor: el intervalo de confianza a 95% es

$$[\bar{y} - 1.96\sigma/\sqrt{n}, \bar{y} + 1.96\sigma/\sqrt{n}]. \qquad \text{C.20}$$

Por tanto, siempre que se conoce σ, se puede decir que se puede construir fácilmente un intervalo de confianza para μ. Al considerar que σ, es desconocida, se debe utilizar una estimación. Sea

$$s = \left(\frac{1}{n-1}\sum_{i=1}^{n}(y_i - \bar{y})^2\right)^{1/2} \qquad \text{C.21}$$

la cual denota la desviación estándar muestral. Después, se obtiene un intervalo de confianza que depende por completo de los datos observados al remplazar σ en la ecuación (C.20) con su estimación, s. Por desgracia, esto no preserva el nivel de confianza a 95% debido a que s depende de la muestra particular. En otras palabras, el intervalo aleatorio $[\bar{Y} \pm 1.96(S/\sqrt{n})]$ ya no contiene a μ con probabilidad de .95 debido a que la constante σ se remplazó con la variable aleatoria S.

¿Cómo se debe proceder? En lugar de utilizar la distribución normal estándar, se debe depender de la distribución t. La distribución t surge del hecho de que

$$\frac{\bar{Y} - \mu}{S/\sqrt{n}} \sim t_{n-1}, \qquad \text{C.22}$$

donde $\bar{Y}$ es el promedio muestral y S es la desviación estándar muestral de la muestra aleatoria $\{Y_1, ..., Y_n\}$. No se probará (C.22); una demostración detallada puede encontrarse en varios libros [por ejemplo, Larsen y Marx (1986, capítulo 7)].

Para construir un intervalo de confianza a 95%, sea c el 97.5-ésimo percentil en la distribución t_{n-1}. En otras palabras, c es el valor tal que 95% del área de la distribución t_{n-1} está entre $-c$ y c: $P(-c < t_{n-1} < c) = .95$. (El valor de c depende de los grados de libertad $n - 1$, aunque no se explicite). La elección de c se ilustra en la figura C.4. Una vez que c se ha elegido de manera apropiada, el intervalo aleatorio $[\bar{Y} - c \cdot S/\sqrt{n}, \bar{Y} + c \cdot S/\sqrt{n}]$ contiene a μ con probabilidad de .95. Para una muestra particular, el intervalo de confianza a 95% se calcula como

$$[\bar{y} - c \cdot s/\sqrt{n}, \bar{y} + c \cdot s/\sqrt{n}]. \qquad \text{C.23}$$

Los valores de c para varios grados de libertad se pueden obtener de la tabla G.2 en el apéndice G. Por ejemplo, si $n = 20$, de manera que los gl son $n - 1 = 19$, entonces $c = 2.093$. Por tanto, el intervalo de confianza a 95 % es $[\bar{y} \pm 2.093(s/\sqrt{20})]$, donde $\bar{y}$ y s son los valores obtenidos de la muestra. Aunque $s = \sigma$ (lo cual es muy improbable), el intervalo de confianza en (C.23) es más amplio que el de (C.20) debido a que $c > 1.96$. Para grados de libertad pequeños, (C.23) es mucho más amplio.

En términos más generales, sea c_α el percentil $100(1 - \alpha)$ en la distribución t_{n-1}. Entonces, se obtiene un intervalo de confianza a $100(1 - \alpha)\%$ como

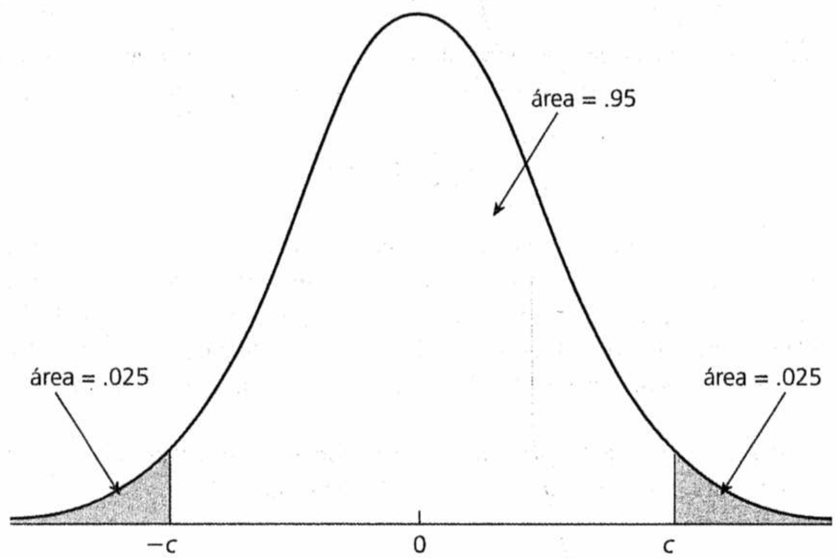
$$[\bar{y} - c_{\alpha/2}s/\sqrt{n}, \bar{y} + c_{\alpha/2}s/\sqrt{n}]. \qquad \text{C.24}$$

Obtener $c_{\alpha/2}$ requiere que se elija α y se conozcan los grados de libertad $n - 1$; entonces, se puede utilizar la tabla G.2. En general, se considerarán los intervalos de confianza a 95%.

Existe una sencilla forma de recordar cómo construir un intervalo de confianza para la media de una distribución normal. Recuerde que $\text{sd}(\bar{Y}) = \sigma/\sqrt{n}$. Por tanto, $s/\sqrt{n}$ es la estimación puntual de $\text{sd}(\bar{Y})$. La variable aleatoria asociada, $S/\sqrt{n}$, algunas veces recibe el nombre de **error**

FIGURA C.4

El 97.5-ésimo percentil, *c*, en una distribución *t*.



estándar de $\bar{Y}$. Debido a la estimación puntual $s/\sqrt{n}$, que se presenta en las fórmulas, se define el error estándar de $\bar{y}$ como $\text{ee}(\bar{y}) = s/\sqrt{n}$. Entonces, (C.24) se puede escribir de forma abreviada como

$$[\bar{y} \pm c_{\alpha/2} \cdot \text{ee}(\bar{y})]. \tag{C.25}$$

Esta ecuación muestra por qué la noción de error estándar de una estimación desempeña una importante función en econometría.

Ejemplo C.2

[Efecto de los subsidios para la capacitación laboral en la productividad de los trabajadores]

Holzer, Block, Cheatham y Knott (1993) estudiaron los efectos de los subsidios para la capacitación laboral en la productividad de los trabajadores al recabar información sobre "tasa de desperdicio" para una muestra de empresas manufactureras de Michigan que recibieron subsidios para la capacitación laboral en 1988. La tabla C.3 lista las tasas de desperdicio, medidas como número de artículos no utilizables y, por tanto, que tenían que desecharse por cada 100 producidos, entre 20 empresas. Cada una de estas empresas recibió un subsidio para la capacitación laboral en 1988; en 1987 no existían los subsidios. Se desea construir un intervalo de confianza para el cambio en la tasa de desperdicio de 1987 a 1988 para la población de todas las empresas manufactureras que pudieran haber recibido subsidios.

Se supone que el cambio en la tasa de desperdicio tiene una distribución normal. Dado que $n = 20$, un intervalo de confianza a 95% para el cambio en la media de las tasas de desperdicio μ es $[\bar{y} \pm 2.093 \cdot \text{ee}(\bar{y})]$, donde $\text{ee}(\bar{y}) = s/\sqrt{n}$. El valor de 2.093 es el 97.5-ésimo percentil en una distribución t_{19}. Para los valores de la muestra en particular, $\bar{y} = -1.15$ y $\text{ee}(\bar{y}) = .54$ (cada uno de ellos redondeado a dos decimales), así que el intervalo de confianza a 95% es $[-2.28, -.02]$. El valor cero se excluye de este intervalo, así que se concluye que, con una confianza a 95%, el cambio promedio en las tasas de desperdicio de la población no es cero.

TABLA C.3

Tasas de desperdicio para 20 empresas manufactureras de Michigan

Empresa	1987	1988	Cambio
1	10	3	–7
2	1	1	0
3	6	5	–1
4	.45	.5	.05
5	1.25	1.54	.29
6	1.3	1.5	.2
7	1.06	.8	–.26
8	3	2	–1
9	8.18	.67	–7.51
10	1.67	1.17	–.5
11	.98	.51	–.47
12	1	.5	–.5
13	.45	.61	.16
14	5.03	6.7	1.67
15	8	4	–4
16	9	7	–2
17	18	19	1
18	.28	.2	–.08
19	7	5	–2
20	3.97	3.83	–.14
Promedio	4.38	3.23	–1.15

En este punto, el ejemplo C.2 es más ilustrativo pues tiene, como un análisis económico, fallas potencialmente más serias. Pero lo más importante es que supone que cualquier reducción sistemática en las tasas de desperdicio se debe a los subsidios en la capacitación laboral. Pero, desde luego, en el transcurso de un año, muchas cosas pueden suceder que modifiquen la productividad del trabajador. A partir de este análisis, no hay forma de saber si la disminución en las tasas de desperdicio se puede atribuir a los subsidios para la capacitación laboral, o si, al menos en parte, es responsable alguna fuerza externa.

Una sencilla regla general para un intervalo de confianza a 95%

El intervalo de confianza en (C.25) se puede calcular para cualquier tamaño muestral y cualquier nivel de confianza. Como se vio en la sección B.5, la distribución t se aproxima a la distribución normal estándar a medida que aumentan los grados de libertad. En particular, para $\alpha = .05$, $c_{\alpha/2} \to 1.96$ conforme $n \to \infty$, aunque $c_{\alpha/2}$ siempre es mayor que 1.96 para cada n. Una *regla general* para un nivel de confianza aproximado a 95% es

$$[\bar{y} \pm 2 \cdot \mathrm{ee}(\bar{y})]. \qquad \boxed{\textbf{C.26}}$$

En otras palabras, se obtiene $\bar{y}$ y su error estándar, y después se calcula $\bar{y}$ menos el doble de su error estándar para obtener el intervalo de confianza. Esto es demasiado amplio para una n, muy grande, y es demasiado reducido para una n pequeña. Como puede verse en el ejemplo C.2, incluso para una n tan pequeña como 20, (C.26) está en el ámbito del intervalo de confianza para la media de una distribución normal. Esto significa que es posible obtener algo muy aproximado a 95% del intervalo de confianza sin tener que referirse a las tablas t.

Intervalos de confianza asintóticos para poblaciones no normales

En algunas aplicaciones, la población es claramente no normal. Un caso sobresaliente es la distribución de Bernoulli, donde la variable aleatoria asume sólo valores de cero y uno. En otros casos, la población no normal no tiene distribución estándar. Esto no importa, siempre y cuando el tamaño de la muestra sea lo bastante grande para el teorema del límite central para dar una buena aproximación de la distribución de la media muestral $\bar{Y}$. Para una n, grande, un intervalo de confianza *aproximado* a 95% es

$$[\bar{y} \pm 1.96 \cdot \mathrm{ee}(\bar{y})], \qquad \boxed{\textbf{C.27}}$$

donde el valor 1.96 es el 97.5-ésimo percentil en la distribución normal estándar. En términos mecánicos, calcular un intervalo de confianza aproximado no difiere del caso normal. Una pequeña diferencia consiste en que el número que multiplica el error estándar proviene de la distribución normal estándar, y no de la distribución t debido a que se están usando asintóticas. Debido a que la distribución t se aproxima a la normal estándar a medida que los gl aumentan, la ecuación (C.25) también se legitima perfectamente como una aproximación al intervalo de 95%; algunos prefieren ésta a (C.27), pues la primera es exacta para poblaciones normales.

Ejemplo C.3

[Discriminación racial en las contrataciones]

El Urban Institute realizó un estudio en 1988 en Washington, D.C. para examinar el grado de discriminación racial en las contrataciones. Se entrevistó a cinco pares de personas para varios puestos. En cada par, una persona era blanca y la otra negra, las cuales ofrecieron currículos que indicaba que eran virtualmente iguales en términos de experiencia, educación y otros factores que determinaba la calificación para el trabajo. La idea era hacer que los individuos fueran tan similares como fuera posible con excepción de la raza. Las personas de cada par pasaban por una entrevista para solicitar el mismo trabajo y los investigadores anotaron qué solicitante recibió cada oferta de trabajo. Este es un ejemplo de un *análisis de pares* *igualados,* donde cada prueba consiste en datos de dos personas (o dos empresas, ciudades, etcétera) que se consideran similares en muchos aspectos pero diferentes en una característica importante.

Sea θ_B la probabilidad de que la persona negra obtenga el trabajo y θ_W la probabilidad de que la persona blanca lo obtenga. Lo que aquí interesa es la diferencia, $\theta_B - \theta_W$. Sea B_i una variable de Bernoulli igual a uno si la persona negra obtiene una oferta de trabajo de su empleador i, y cero de otra manera. Asimismo, $W_i = 1$ si la persona blanca obtiene una oferta de trabajo del empleador i, y cero en caso contrario. Al reunir los resultados de los cinco pares de personas, hubo un total de $n = 241$ pruebas (pares de entrevistas con empleadores). Los estimadores insesgados de θ_B y θ_W son $\bar{B}$ y $\bar{W}$, las fracciones de entre-vistas en las cuales personas blancas y negras recibieron ofertas de trabajo, respectivamente.

Para calcular un intervalo de confianza para la media de una población, defina una nueva variable $Y_i = B_i - W_i$. Ahora, Y_i puede tomar tres valores: -1 si la persona negra no obtiene el trabajo, pero la persona blanca sí, 0 si ninguna de las personas obtiene el trabajo o si las dos lo obtienen, y 1 si la persona negra obtiene el trabajo y la persona blanca no. Entonces, $\mu \equiv \mathrm{E}(Y_i) = \mathrm{E}(B_i) - \mathrm{E}(W_i) = \theta_B - \theta_W$.

La distribución de Y_i sin lugar a dudas es no normal; es discreta y sólo toma tres valores. Sin embargo, un intervalo de confianza aproximado para $\theta_B - \theta_W$ se puede obtener mediante métodos muestrales grandes.

Mediante los 241 puntos de datos observados $\bar{b} = .224$ y $\bar{w} = .357$, así que $\bar{y} = .224 - .357 = -.133$. Por tanto, 22.4% de los solicitantes negros recibieron oferta de trabajo, mientras que 35.7% de los solicitantes blancos recibieron oferta de trabajo. Esto es a *primera vista* una evidencia de la discriminación en contra de las personas negras, pero es posible saber mucho más al calcular un intervalo de confianza para μ. Para calcular un intervalo de confianza a 95%, se necesita la desviación estándar muestral. Ésta resulta ser $s = .482$ [mediante la ecuación (C.21)]. Mediante (C.27) se obtiene un IC a 95% para $\mu = \theta_B - \theta_W$ cuando $-.133 \pm 1.96(.482/\sqrt{241}) = -.133 \pm .031 = [-.164, -.102]$. El IC aproximado a 99% es $-.133 \pm 2.58(.482/\sqrt{241}) = [-.213, -.053]$. Naturalmente, esto contiene a un rango más amplio de valores que el IC a 95%. Pero incluso el IC a 99% no contiene al valor cero. Por tanto, se tiene un alto grado de confianza en que la diferencia poblacional $\theta_B - \theta_W$ no sea cero.

Antes de regresar a la prueba de hipótesis, es útil revisar las diferentes cantidades muestrales y poblacionales que miden las diferencias entre las distribuciones poblacionales y las distribuciones de muestreo de los estimadores. Estas cantidades suelen aparecer en los análisis estadísticos y sus extensiones son importantes para el análisis de regresión en el texto principal. La cantidad σ es la desviación estándar de la población (desconocida); es la medida de la diferencia en la distribución de Y. Cuando se divide σ entre $\sqrt{n}$, se obtiene la **desviación estándar muestral** de $\bar{Y}$ (el promedio muestral). Si bien σ es una característica fija de la población, $\mathrm{sd}(\bar{Y}) = \sigma/\sqrt{n}$ se reduce a cero conforme $n \rightarrow \infty$: el estimador de μ se vuelve cada vez más preciso a medida que el tamaño de la muestra aumenta.

La estimación de σ para una muestra en particular, s, recibe el nombre de desviación estándar muestral debido a que se obtiene de la muestra. (También recibe el nombre de variable aleatoria subyacente, S, que cambia a través de muestras diferentes, la desviación estándar muestral.) Al igual que $\bar{y}$ como estimación de μ, s es el "mejor supuesto" para σ dada la muestra en cuestión. La cantidad $s/\sqrt{n}$ es lo que se llama error estándar de $\bar{y}$, y es la mejor estimación de $\sigma/\sqrt{n}$. Los intervalos de confianza para el parámetro poblacional μ dependen directamente de $\mathrm{ee}(\bar{y}) = s/\sqrt{n}$. Debido a que este error estándar se reduce a cero a medida que el tamaño de la muestra aumenta, un tamaño de muestra más grande, por lo general, significa un intervalo de confianza menor. Por tanto, se puede ver claramente que un beneficio de tener más datos es que generan intervalos de confianza más estrechos. La noción de error estándar de una estimación, que en la vasta mayoría de los casos se reduce a cero a una tasa de $1/\sqrt{n}$, desempeña una función fundamental en la prueba de hipótesis (como se verá en la siguiente sección) y para intervalos de confianza y prueba en el contexto de regresión múltiple (como se analizó en el capítulo 4).

C.6 Prueba de hipótesis

Hasta ahora se ha analizado cómo evaluar los estimadores puntuales, y se ha visto, en el caso de una media poblacional, cómo construir e interpretar intervalos de confianza. Pero en ocasiones la cuestión que interesa es tener una respuesta definitiva de sí o no. He aquí algunos ejemplos: 1) ¿Un programa de capacitación laboral aumenta efectivamente la productividad promedio del trabajador? (vea el ejemplo C.2); 2) ¿Las personas negras sufren discriminación en el proceso de contratación? (vea el ejemplo C.3); 3) ¿Las estrictas leyes estatales contra la conducción en estado de ebriedad reducen el número de arrestos por conducir en esas condiciones? Diseñar métodos para contestar tales preguntas, mediante una muestra de datos, recibe el nombre de prueba de hipótesis.

Fundamentos de la prueba de hipótesis

Para ilustrar las cuestiones que implica la prueba de hipótesis, considere un ejemplo electoral. Suponga que se postulan dos candidatos en una elección, el candidato A y el B. El candidato A reportó haber recibido 42% del voto popular, mientras que el candidato B recibió 58%. Se supone que estos son porcentajes verdaderos entre la población votante y se tratarán como tales.

El candidato A está convencido de que más personas votaron por él, así que le gustaría investigar si la elección fue fraudulenta. Con cierto conocimiento estadístico, el candidato A contrata a una agencia consultora para hacer una muestra al azar de 100 votantes para saber si cada persona votó o no por él. Suponga que, en la muestra recabada, 53 personas votaron por el candidato A. Esta estimación muestral de 53% excede claramente el valor reportado de la población de 42%. ¿El candidato A debe concluir que en realidad la elección fue fraudulenta?

Si bien todo parece indicar que hubo una omisión de votos en el conteo para el candidato A, es posible que, en una muestra de 100, se observen 53 personas que en realidad votaron por el candidato A. La pregunta es: ¿Qué tan *sólida* es la evidencia muestral frente al porcentaje oficialmente reportado de 42%?

Una forma de proceder es establecer una **prueba de hipótesis**. Sea θ la verdadera proporción de la población que votó por el candidato A. La hipótesis de que los resultados reportados son precisos se puede expresar como

$$H_0: \theta = .42. \qquad \text{(C.28)}$$

Éste es un ejemplo de **hipótesis nula**. Siempre se denotará aquí la hipótesis nula como H_0. En la prueba de hipótesis, la hipótesis nula desempeña una función similar a la de un defensor de oficio en muchos sistemas judiciales; tal como el acusado se presume inocente hasta demostrarse culpable, la hipótesis nula se presume verdadera hasta que los datos sugieran de manera contundente lo contrario. En este ejemplo, el candidato A debe presentar evidencia concluyente en contra de (C.28) con el fin de ganar el derecho a un recuento.

La **hipótesis alternativa** en el ejemplo de la elección es que la verdadera proporción que votó por el candidato A en la elección es mayor que .42:

$$H_1: \theta > .42. \qquad \text{(C.29)}$$

Con el fin de concluir que H_0 es falsa y H_1 es verdadera, se debe tener la evidencia "más allá de la duda razonable" en contra de H_0. ¿Cuántos votos de 100 serían necesarios antes de parecer que la evidencia es contundente en contra de H_0? La mayoría estaría de acuerdo en que 43 votos de una muestra de 100 no son suficientes para anular los resultados de la elección original; tal resultado está dentro de la variación muestral esperada. Por otra parte, no es necesario observar 100 votos para el candidato A para que se dude de H_0. Si 53 de 100 basta para rechazar H_0 es mucho menos claro. La respuesta depende de cómo se cuantifique la expresión "más allá de la duda razonable".

Antes de regresar a la cuestión de cuantificar la incertidumbre en la prueba de hipótesis, es necesario prevenir cualquier posible confusión. Quizá se haya observado que la hipótesis en las ecuaciones (C.28) y (C.29) no agota todas las posibilidades: podría ser que θ sea menor que .42. Para la aplicación en cuestión, esa posibilidad no es de particular interés; no tiene nada que ver con anular los resultados de la elección. Por tanto, sólo se puede afirmar al principio que se están ignorando las alternativas θ con $\theta < .42$. Sin embargo, algunos autores prefieren establecer hipótesis nulas y alternativas, con fines de exhaustividad, en cuyo caso la hipótesis nula aquí sería H_0: $\theta \leq .42$. Expresado de esta forma, la hipótesis nula es una hipótesis nula *compuesta*, debido a que permite más de un valor bajo H_0. [Por el contrario, la ecuación (C.28) es un ejemplo de una hipótesis nula *simple*]. Para este tipo de ejemplos, no importa si se establece la hipótesis nula como en (C.28) o como una nula compuesta: el valor más difícil de rechazar si $\theta \leq .42$ es $\theta = .42$. (Es decir, si se rechaza el valor $\theta = .42$ frente a $\theta > .42$, entonces lógicamente se deberá rechazar cualquier valor menor que .42.) Por tanto, este procedimiento de prueba basado en (C.28) lleva a la misma prueba como si H_0: $\theta \leq .42$. En este libro siempre se establecerá una hipótesis nula como una hipótesis nula simple.

En la prueba de hipótesis se pueden cometer dos tipos de errores. Primero, se puede rechazar la hipótesis nula cuando de hecho es verdadera. Esto recibe el nombre de **error tipo I**. En el ejemplo de la elección ocurre un error tipo I si se rechaza H_0 cuando la verdadera proporción de personas que votan por el candidato A es en realidad .42. El segundo tipo de error es no rechazar H_0 cuando en realidad es falsa. Esto recibe el nombre de **error tipo II**. En el ejemplo de la elección, ocurre un error tipo II si $\theta > .42$ pero no se rechaza H_0.

Después de que haber tomado la decisión de rechazar o no la hipótesis nula, se habrá decidido correctamente o se habrá cometido un error. Nunca se sabrá con certeza si se cometió un error. No obstante, se puede calcular la *probabilidad* de cometer un error tipo I o tipo II. Las reglas de prueba de hipótesis se construyen para hacer que la probabilidad de cometer un error tipo I sea muy pequeña. En general, se define el **nivel de significancia** (o simplemente el *nivel*) de una prueba como la probabilidad de un error tipo I; suele denotarse por α. En símbolos, se tiene

$$\alpha = \mathrm{P}(\text{Reject } H_0|H_0). \tag{C.30}$$

El lado derecho se lee como: "La probabilidad de rechazar H_0 dado que H_0 es verdadera."

La prueba clásica de la hipótesis requiere que inicialmente se especifique un nivel de significancia para una prueba. Cuando se especifica un valor para α, esencialmente se está cuantificando la tolerancia para un error tipo I. Los valores comunes de α son .10, .05 y .01. Si $\alpha = .05$, entonces el investigador desea rechazar falsamente H_0 5% de las veces, con el fin de detectar desviaciones de H_0.

Una vez que se ha elegido el nivel de significancia, entonces quizá se desee minimizar la probabilidad del error tipo II. Otra posibilidad es que se quiera maximizar la **potencia de una prueba** contra cualquier alternativa relevante. La potencia de una prueba es sólo uno menos la probabilidad del error tipo II. Matemáticamente,

$$\pi(\theta) = \mathrm{P}(\text{Reject } H_0|\theta) = 1 - \mathrm{P}(\text{Type II}|\theta),$$

donde θ denota el valor real del parámetro. Naturalmente, se desearía poder igualar a uno si la hipótesis nula es falsa. Pero eso es imposible de lograr mientras se mantiene pequeño el nivel de significancia. En cambio, se opta porque las pruebas maximicen la potencia para un nivel de significancia dado.

Pruebas de hipótesis para la media de una población normal

Con el fin de probar una hipótesis nula contra una alternativa, es necesario elegir un estadístico de prueba (o estadístico, para abreviar) y un valor crítico. La elección del estadístico y el valor crítico está basada en la conveniencia y el deseo de maximizar la potencia dado un nivel de significancia para la prueba. En esta subsección se revisará cómo comprobar las hipótesis para la media de una población normal.

Un **estadístico de prueba**, denotado como T, es alguna función de la muestra aleatoria. Cuando se calcula el estadístico para un resultado en particular, se obtiene un resultado para el estadístico de prueba, que se denotará por t.

Dado un estadístico de prueba, se puede definir una regla de rechazo que determine cuando H_0 se rechace en favor de H_1. En esta prueba todas las reglas de rechazo están basadas en comparar el valor de un estadístico de prueba, t, respecto a un **valor crítico**, c. Los valores de t que resulten del rechazo de la hipótesis nula se conocen de manera conjunta como la **región de rechazo**. Para determinar el valor crítico, se debe decidir primero el nivel de significancia de la prueba. Después, dada α, el valor crítico asociado con α se determina mediante la distribución de T, *suponiendo* que H_0 sea verdadera. Se escribirá este valor crítico como c y se eliminará el hecho de que depende de α.

Probar la hipótesis acerca de la media μ de una población Normal(μ,σ^2) es sencillo. La hipótesis nula se expresa como

$$H_0\colon \mu = \mu_0, \tag{C.31}$$

donde μ_0 es un valor que se especifica. En la mayoría de las aplicaciones, $\mu_0 = 0$, pero el caso general no es más difícil.

La regla de rechazo que se eligió depende de la naturaleza de la hipótesis alternativa. Las tres alternativas de interés son

$$H_1\colon \mu > \mu_0, \tag{C.32}$$

$$H_1\colon \mu < \mu_0, \tag{C.33}$$

y

$$H_1\colon \mu \neq \mu_0. \tag{C.34}$$

La ecuación (C.32) da una **alternativa de una cola**, como (C.33). Cuando la hipótesis alternativa es (C.32), la nula es efectivamente $H_0\colon \mu \leq \mu_0$, dado que se rechaza H_0 sólo cuando $\mu > \mu_0$. Esto es apropiado cuando se tiene interés en el valor de μ sólo cuando μ es al menos tan grande como μ_0. La ecuación (C.34) es una **alternativa de dos colas**. Esta es apropiada cuando se tiene interés en cualquier desviación a partir de la hipótesis nula.

Considere primero la alternativa en (C.32). Intuitivamente, se debe rechazar H_0 en favor de H_1 cuando el valor del promedio muestral, $\bar{y}$, es lo "bastante" mayor que μ_0. Pero, ¿cómo se debe determinar cuando $\bar{y}$ es lo bastante mayor para que H_0 se rechace al nivel de significancia elegido? Esto requiere conocer la probabilidad de rechazar la hipótesis nula, cuando ésta es verdadera. En lugar de trabajar directamente con $\bar{y}$, se utiliza su versión estandarizada, donde σ se remplaza con la desviación estándar muestral, s:

$$t = \sqrt{n}(\bar{y} - \mu_0)/s = (\bar{y} - \mu_0)/\mathrm{ee}(\bar{y}), \tag{C.35}$$

donde $\text{ee}(\bar{y}) = s/\sqrt{n}$ es el error estándar de $\bar{y}$. Dada la muestra de datos, es fácil obtener t. Se trabaja con t debido a que, bajo la hipótesis nula, la variable aleatoria

$$T = \sqrt{n}(\bar{Y} - \mu_0)/S$$

tiene una distribución t_{n-1}. Ahora, suponga que se ha establecido un nivel de significancia a 5%. Entonces se elije el valor crítico c, de manera que $P(T > c|H_0) = .05$; es decir, la probabilidad de un error tipo I es de 5%. Una vez que se ha encontrado c, la regla de rechazo es

$$t > c, \qquad \boxed{\textbf{C.36}}$$

donde c es el percentil $100(1 - \alpha)$ en una distribución t_{n-1}; como porcentaje, el nivel de significancia es $100 \cdot \alpha\%$. Este es un ejemplo de una **prueba de una cola** debido a que la región de rechazo está en una cola de la distribución t. Para un nivel de significancia de 5%, c es el percentil 95-énesimo en la distribución t_{n-1}; esto se ilustra en la figura C.5. Un nivel de significancia diferente ocasiona un valor crítico diferente.

El estadístico en la ecuación (C.35) suele recibir el nombre de **estadístico t** para prueba de H_0: $\mu = \mu_0$. El estadístico t mide la distancia de $\bar{y}$ a μ_0 *con relación* al error estándar de $\bar{y}$, $\text{ee}(\bar{y})$.

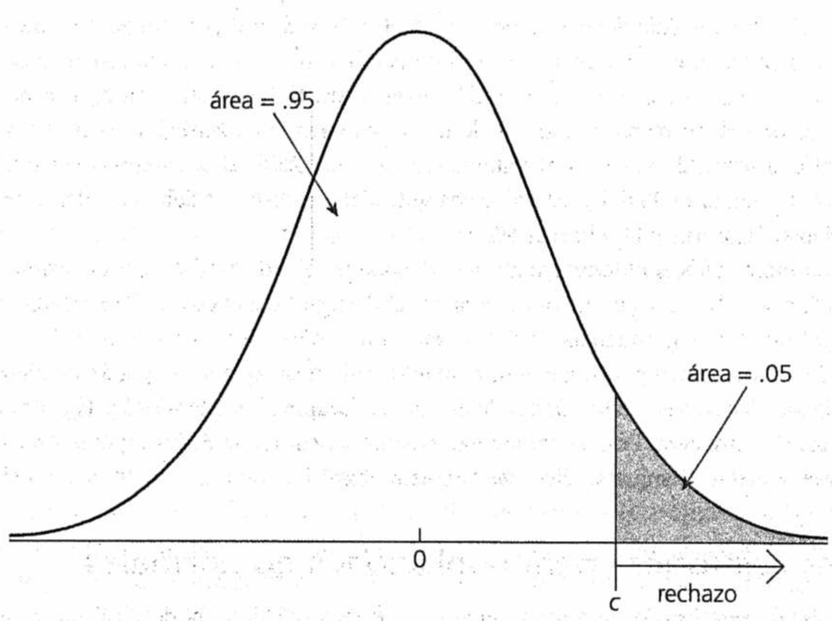
Ejemplo C.4

[Efecto de las zonas empresariales en las inversiones de negocios]

En la población de ciudades que otorgan zonas empresariales en un estado en particular [vea Papke (1994) para Indiana], sea Y el cambio porcentual en la inversión a partir del año anterior al año en que una ciudad se

FIGURA C.5

Región de rechazo para una prueba con un nivel de significancia de 5% frente a la alternativa de una cola $\mu > \mu_0$.



convirtió en una zona empresarial. Suponga que Y tiene una distribución Normal(μ,σ^2). La hipótesis nula de que las zonas empresariales no tienen efecto en la inversión de negocios es H_0: $\mu = 0$; la alternativa de que tiene un efecto positivo es H_1: $\mu > 0$. (Se supone que no tienen un efecto negativo). Suponga que se desea probar H_0 al nivel de 5%. El estadístico de prueba en este caso es

$$t = \frac{\bar{y}}{s/\sqrt{n}} = \frac{\bar{y}}{\mathrm{ee}(\bar{y})}. \qquad \boxed{\text{C.37}}$$

Suponga que se tiene una muestra de 36 ciudades a las que les han otorgado zonas empresariales. Entonces, el valor crítico es $c = 1.69$ (vea la tabla G.2) y se rechaza H_0 en favor de H_1 si $t > 1.69$. Suponga que la muestra produce $\bar{y} = 8.2$ y $s = 23.9$. Entonces, $t \approx 2.06$, y por tanto H_0 se rechaza al nivel de 5%. Por consiguiente, se concluye que, al nivel de significancia de 5%, las zonas empresariales tienen un efecto en la inversión promedio. El valor crítico de 1% es 2.44, así que H_0 no se rechaza al nivel de 1%. La misma advertencia se mantiene aquí como en el ejemplo C.2: no se controlan otros factores que pudieran afectar la inversión en las ciudades a través del tiempo, así que no se puede afirmar que el efecto sea causal.

La regla de rechazo es similar a la alternativa de una cola (C.33). Una prueba con un nivel de significancia de $100\cdot\alpha\%$ rechaza H_0 frente a (C.33) siempre que

$$t < -c; \qquad \boxed{\text{C.38}}$$

en otras palabras, se está buscando valores negativos del estadístico t lo cual implica $\bar{y} < \mu_0$ que están lo bastante lejos de cero para rechazar H_0.

Para alternativas de dos colas, se debe tener cuidado en elegir el valor crítico de manera que el nivel de significancia de la prueba siga siendo α. Si H_1 está dada por H_1: $\mu \neq \mu_0$, entonces se rechaza H_0 si $\bar{y}$ está lejos de μ_0 en *valor absoluto*: una $\bar{y}$ mucho mayor o mucho menor que μ_0 ofrece evidencia contra H_0 en favor de H_1. Una prueba al nivel $100\cdot\alpha\%$ se obtiene de la regla de rechazo

$$|t| > c, \qquad \boxed{\text{C.39}}$$

donde $|t|$ es el valor absoluto del estadístico t en (C.35). Esto da una **prueba de dos colas**. Se debe tener cuidado en elegir el valor crítico: c es el percentil $100(1 - \alpha/2)$ en la distribución t_{n-1}. Por ejemplo, si $\alpha = .05$, entonces el valor crítico es el percentil 97.5-ésimo en la distribución t_{n-1}. Esto asegura que H_0 se rechace sólo 5% de las veces cuando es verdadera (vea la figura C.6). Por ejemplo, si $n = 22$, entonces el valor crítico es $c = 2.08$, el 97.5-ésimo percentil en una distribución t_{21} vea la tabla G.2). El valor absoluto del estadístico t debe exceder a 2.08 con el fin de rechazar H_0 frente a H_1 al nivel 5%.

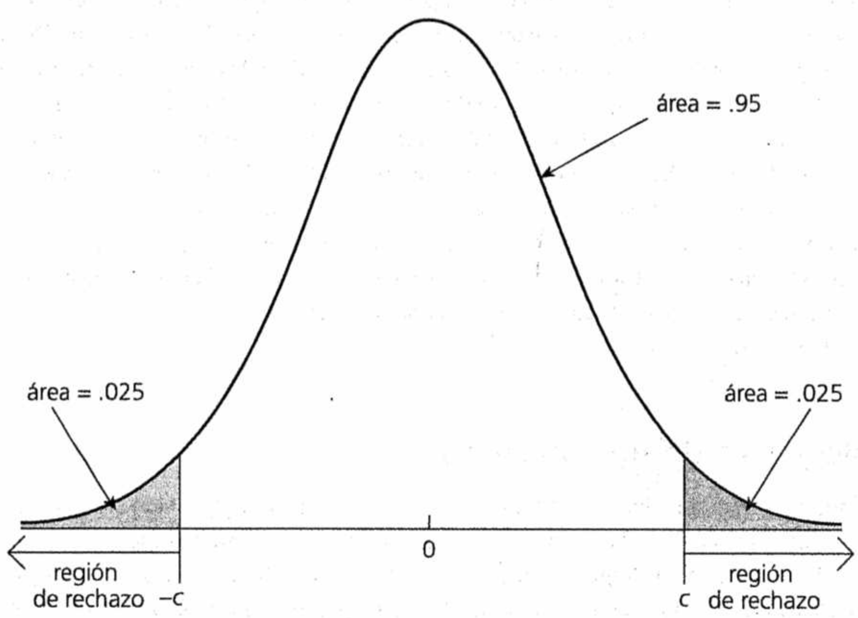
Es importante conocer el lenguaje apropiado en la prueba de hipótesis. En ocasiones la frase apropiada "no se rechazó H_0 en favor de H_1 al nivel de significancia de 5%" se remplaza con "se acepta H_0 al nivel de significancia de 5%." Esta última redacción es incorrecta. Con el mismo conjunto de datos, existen por lo general, numerosas hipótesis que no se pueden rechazar. En el ejemplo de las elecciones, sería ilógico decir que se "aceptan" H_0: $\theta = .42$ y H_0: $\theta = .43$ dado que sólo una es verdadera. Pero es totalmente posible que ninguna de las hipótesis pueda rechazarse. Por esta razón, siempre se dice que "no se rechazó H_0" en lugar de "se aceptó H_0."

Pruebas asintóticas para poblaciones no normales

Si el tamaño muestral es lo bastante grande para invocar el teorema del límite central (vea la sección C.3) la mecánica de la prueba de hipótesis para las medias poblacionales es la *misma*,

FIGURA C.6

Región de rechazo para una prueba al nivel de significancia de 5% frente a la alternativa de dos colas H_1: $\mu \neq \mu_0$.



sin importar si la distribución poblacional es normal o no. La justificación teórica proviene del hecho que, bajo la hipótesis nula

$$T = \sqrt{n}(\bar{Y} - \mu_0)/S \stackrel{a}{\sim} \text{Normal}(0,1).$$

Por tanto, con n, grande, se puede comparar el estadístico t en (C.35) con los valores críticos de una distribución normal estándar. Debido a que la distribución t_{n-1} converge a la distribución normal estándar a medida que n aumenta, t y los valores críticos normales estándar estarán muy cerca para una n extremadamente grande. Debido a que la teoría asintótica está basada en una n que crece de manera ilimitada, no puede indicar si son mejores los valores críticos normales estándares o los t. Para los valores moderados de n, por ejemplo, entre 30 y 60, es tradicional emplear la distribución t debido a que se sabe que es lo correcto para las poblaciones normales. Para $n > 120$, la elección entre t y las distribuciones normales estándar suele ser irrelevante debido a que los valores críticos prácticamente son los mismos.

Debido a que los valores críticos elegidos utilizan la distribución t o la normal estándar son sólo aproximadamente válidos para poblaciones no normales, los niveles de significancia elegidos son sólo aproximados; por tanto, para poblaciones no normales, los niveles de significancia son realmente niveles de significancia asintóticos. De este modelo, si se elige un nivel de significancia de 5%, pero la población es no normal, entonces el nivel de significancia real será mayor o menor que 5% (y no se puede saber cuál es el caso). Cuando el tamaño muestral es grande, el nivel de significancia real será muy cercano a 5%. En términos prácticos, la distinción no es importante, por tanto, se elimina el calificativo "asintótico".

Ejemplo C.5

[Discriminación racial en las contrataciones]

En el estudio del Urban Institute sobre discriminación en la contratación (vea el ejemplo C.3), el interés principal estribaba en probar H_0: $\mu = 0$ frente a H_1: $\mu < 0$, donde $\mu = \theta_B - \theta_W$ es la diferencia de probabilidades de que las personas negras recibieran ofertas de trabajo con relación a los blancos. Recuerde que μ es la media poblacional de la variable $Y = B - W$, donde B y W son los indicadores binarios. Mediante las $n = 241$ comparaciones pareadas, se obtuvo $\bar{y} = -.133$ y $\text{ee}(\bar{y}) = .482/\sqrt{241} \approx .031$. El estadístico t para probar H_0: $\mu = 0$ es $t = -.133/.031 \approx -4.29$. Se recordará del apéndice B que la distribución normal estándar es, para fines prácticos, indistinguible de la distribución t con 240 grados de libertad. El valor de -4.29 está tan lejano en el extremo izquierdo de la distribución que se rechaza H_0 a cualquier nivel de significancia razonable. De hecho, el valor crítico de .005 (medio punto porcentual para una prueba de una cola) es de aproximadamente -2.58. Un valor t de -4.29 es una evidencia *muy* sólida contra H_0 y en favor de H_1. Por tanto, se concluye que existe discriminación en la contratación.

Cálculo y uso de los valores-*p*

El requerimiento tradicional de usar un nivel de significancia por adelantado significa que investigadores diferentes, con ayuda de los mismos datos y el mismo procedimiento para comprobar la misma hipótesis, pueden llegar a conclusiones diferentes. Reportar el nivel de significancia al cual se lleva a cabo la prueba resuelve hasta cierto punto este problema, pero no lo elimina por completo.

Para ofrecer más información, se puede plantear la siguiente pregunta: ¿cuál es el nivel de significancia *máximo* al cual se puede realizar la prueba y aun así no rechazar la hipótesis nula? El valor se conoce como el **valor-*p*** de una prueba (en ocasiones llamado el valor de la prob). En comparación con la elección de un nivel de significancia por adelantado y obtener un valor crítico, comparar un valor-p es un tanto más difícil. Pero con la ayuda de los cálculos computacionales rápidos y económicos, los valores-p ahora son muy fáciles de obtener.

Como ejemplo, considere el problema de probar H_0: $\mu = 0$ en una población Normal(μ,σ^2) El estadístico t en este caso es $T = \sqrt{n}\cdot\bar{Y}/S$, y se supone que n es suficientemente grande para tratar T como si contara con una distribución normal estándar según H_0. Suponga que el valor observado de T para la muestra es $t = 1.52$. (Observe cómo se ha saltado el paso de elegir un nivel de significancia). Ahora que se ha visto el valor t, se encuentra el nivel de significancia más grande al cual no se lograría rechazar H_0. Este es el nivel de significancia asociado con usar t como valor crítico. Debido a que la prueba del estadístico T tiene una distribución normal estándar según H_0, se tiene

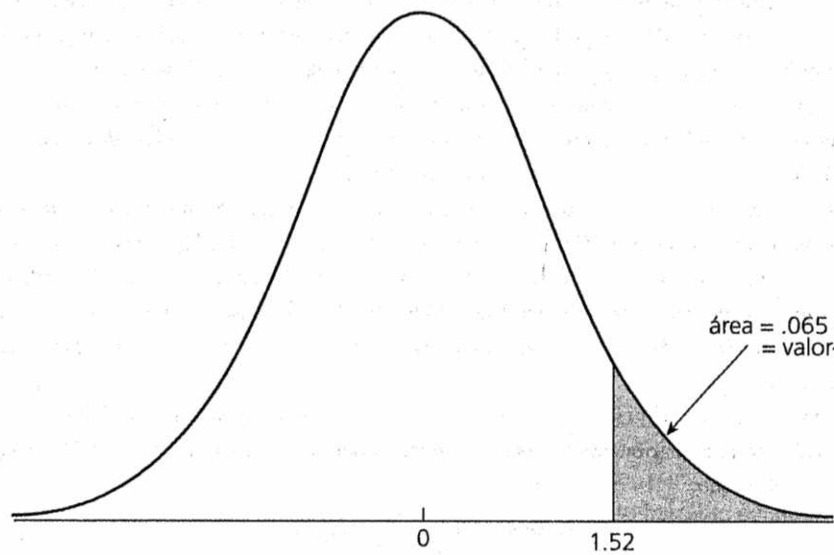
$$\text{valor-}p = P(T > 1.52|H_0) = 1 - \Phi(1.52) = .065, \qquad \textbf{C.40}$$

donde $\Phi(\cdot)$ denota la fda normal estándar. En otras palabras el valor-p en este ejemplo es simplemente el área a la derecha de 1.52, el valor observado del estadístico de prueba, en una distribución normal estándar. Vea la figura C.7 como ejemplo.

Debido a que el valor-p = .065, el nivel de significancia mayor al cual se puede realizar esta prueba y no rechazar es 6.5%. Si se realiza la prueba al nivel por debajo de 6.5% (como a 5%), no se rechaza H_0. Si se realiza la prueba a un nivel mayor que 6.5% (como a 10%), se rechaza H_0. Con el valor-p a la mano, se puede realizar la prueba a cualquier nivel.

FIGURA C.7

El valor-p cuando $t = 1.52$ para una alternativa de una cola $\mu > \mu_0$.



El valor-p en este ejemplo tiene otra interpretación útil: es la probabilidad que observe un valor de T tan grande como 1.52 cuando la hipótesis nula es verdadera. Si la hipótesis nula en realidad es verdadera, se observaría un valor de T tan grande como 1.52 debido a la probabilidad de sólo 6.5% de las veces. Si esto es lo bastante pequeño para rechazar H_0 depende de la tolerancia que se tenga para un error tipo I. El valor-p tiene una interpretación similar en todos los demás casos, como se verá.

En general, los valores-p pequeños son evidencia *contra* H_0, dado que indican que el resultado de los datos ocurre con una probabilidad pequeña si H_0 es verdadera. En el ejemplo anterior, si t hubiera sido un valor mayor, por ejemplo, $t = 2.85$, entonces el valor-p habría sido $1 - \Phi(2.85) \approx .002$. Esto significa que, si la hipótesis nula fuera verdadera, se observaría un valor de T hasta de 2.85 con una probabilidad de .002. ¿Cómo interpretar esto? Ya sea que se obtenga una muestra muy inusual o que la hipótesis nula sea falsa. A menos que se tenga una tolerancia *muy* pequeña para un error de tipo I, se rechazaría la hipótesis nula. Por otra parte, un valor-p grande es una evidencia débil contra H_0. Si se obtuvo $t = .47$ en el ejemplo anterior, entonces el valor-$p = 1 - \Phi(.47) = .32$. Observar a un valor de T mayor que .47 sucede con probabilidad de .32 incluso cuando H_0 es verdadera; esto es lo bastante grande de manera que la duda acerca de H_0 es insuficiente, a menos que se tenga una tolerancia muy alta para un error tipo I.

Para probar hipótesis acerca de la media de una población mediante la distribución t, se necesitan tablas detalladas para calcular los valores-p. La tabla G.2 sólo permite poner límites a los valores-p. Por fortuna, numerosos paquetes estadísticos y econométricos ahora calculan los valores-p de manera rutinaria, y también ofrecen cálculos de las fda para t y otras distribuciones usadas para calcular valores-p.

Ejemplo C.6

[Efecto de los subsidios para la capacitación laboral en la productividad de los trabajadores]

Considere nuevamente los datos de Holzer *et al.* (1993) en el ejemplo C.2. Desde una perspectiva de políticas, existen dos cuestiones de interés. Primero, ¿cuál es la mejor estimación del cambio promedio en las tasas de desperdicios industriales, μ? Ya se obtuvo esto de la muestra de 20 empresas listadas en la tabla C.3: el promedio muestral en el cambio en las tasas de desperdicio es de -1.15. Con relación a la tasa inicial de desperdicio en 1987, esto representa una disminución en la tasa de desperdicio de aproximadamente 26.3% ($-1.15/4.38 \approx -.263$), lo cual es un efecto importante.

También se desea saber si la muestra ofrece una evidencia contundente para un efecto en la población de empresas manufactureras que pudieran haber recibido los subsidios. La hipótesis nula es H_0: $\mu = 0$, y se prueba esto contra H_1: $\mu < 0$, donde μ es el cambio promedio en las tasas de desperdicio. Según la nula, los subsidios de capacitación laboral no tienen efecto en las tasas promedio de desperdicio. La alternativa establece que existe un efecto. La alternativa $\mu > 0$ no es de interés; la hipótesis nula es efectivamente H_0: $\mu \geq 0$.

Dado que $\bar{y} = -1.15$ y $\text{ee}(\bar{y}) = .54$, $t = -1.15/.54 = -2.13$. Esto se halla por debajo del valor crítico de 5% de -1.73 (de una distribución t_{19}) pero por encima del valor crítico de 1%, -2.54. El valor-p en este caso se calcula como

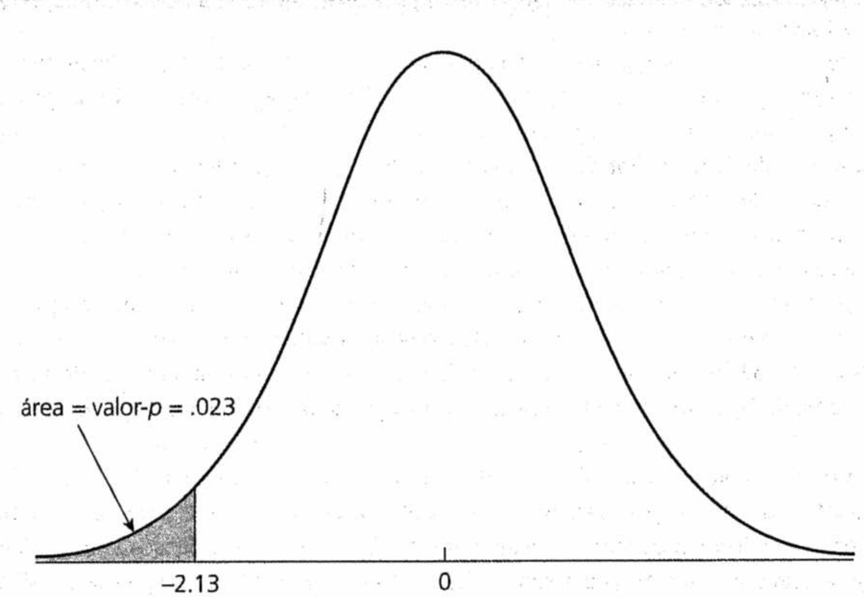
$$\text{valor-}p = \text{P}(T_{19} < -2.13), \tag{C.41}$$

donde T_{19} representa una variable aleatoria que se distribuye como una t con 19 grados de libertad. La desigualdad se invierte de (C.40) debido a que la alternativa tiene la forma en (C.33). La probabilidad en (C.41) es el área a la izquierda de -2.13 en una distribución t_{19} (vea la figura C.8).

Mediante la tabla G.2, lo más que se puede decir es que el valor-p está entre .025 y .01, pero esto es más cercano a .025 (dado que el 97.5-ésimo percentil es de aproximadamente 2.09). Mediante un software

FIGURA C.8

El valor-p cuando $t = -2.13$ con 19 grados de libertad para una alternativa de una cola $\mu < 0$.



estadístico, como Stata, se puede calcular el valor-p exacto. Resulta ser de aproximadamente .023, que es una evidencia razonable contra H_0. Ésta es ciertamente suficiente evidencia para rechazar la hipótesis nula de que los subsidios para la capacitación no tienen efecto al nivel de significancia de 2.5% (y por tanto, al nivel de 5%).

Calcular un valor-p para una prueba de dos colas es similar, pero se debe dar cuenta de la naturaleza de dos colas de la regla de rechazo. Para la prueba t en torno a la media poblacional, el valor-p se calcula como

$$P(|T_{n-1}| > |t|) = 2P(T_{n-1} > |t|), \qquad \boxed{\text{C.42}}$$

donde t es el valor del estadístico de prueba y T_{n-1} es una variable aleatoria t. (Para una n grande, remplazar T_{n-1} con una variable aleatoria estándar normal). Por tanto, se calcula el valor absoluto del estadístico t, se encuentra el área a la derecha de este valor en una distribución t_{n-1} y se multiplica el área por dos.

Para poblaciones no normales, el valor-p exacto puede ser difícil de obtener. Sin embargo, se pueden encontrar valores-p *asintóticos* al utilizar estos mismos cálculos. Estos valores-p son válidos para tamaños muestrales grandes. Para una n mayor que, por ejemplo, 120, se puede emplear también la distribución normal estándar. La tabla G.1 es lo bastante detallada para obtener valores-p precisos, pero también se puede usar un programa estadístico o de econometría.

Ejemplo C.7

[Discriminación racial en las contrataciones]

Mediante los datos pareados del Urban Institute ($n = 241$), se obtuvo $t = -4.29$. Si Z es una variable aleatoria normal estándar, $P(Z < -4.29)$ es, para fines prácticos, cero. En otras palabras, el valor-p asintótico para este ejemplo es esencialmente cero. Esta es una evidencia contundente contra H_0.

Resumen de cómo usar los valores-*p*:

i) Elija un estadístico T de prueba y decida la naturaleza de la alternativa. Esto determina si la regla de rechazo es $t > c$, $t < -c$, o $|t| > c$.

ii) Use el valor observado del estadístico t como el valor crítico y calcule el nivel de significancia correspondiente de la prueba. Este es el valor-p. Si la regla de rechazo es de la forma $t > c$, entonces el valor-$p = P(T > t)$. Si la regla de rechazo es $t < -c$, entonces el valor-$p = P(T < t)$; si la regla de rechazo es $|t| > c$, entonces el valor-$p = P(|T| > |t|)$.

iii) Si se eligió un nivel de significancia α entonces se rechaza H_0 al nivel $100\cdot\alpha\%$ si el valor-$p < \alpha$. Si el valor-$p \geq \alpha$, entonces no se rechaza H_0 al nivel $100\cdot\alpha\%$. Por tanto, este es un valor-p pequeño que lleva al rechazo.

La relación entre intervalos de confianza y pruebas de hipótesis

Debido a que construir intervalos de confianza y pruebas de hipótesis implican enunciados de probabilidad, es natural pensar que existe una conexión entre ellos. Resulta que así es. Después de que se construye el intervalo de confianza, se pueden realizar varias pruebas de hipótesis.

Los intervalos de confianza que se han analizado son todos de naturaleza de dos colas. (En este libro, no habrá necesidad de construir intervalos de confianza de una cola). Por tanto, los intervalos de confianza se pueden usar para probar contra alternativas de *dos colas*. En el caso de una media poblacional, la nula está dada por (C.31) y la alternativa por (C.34). Suponga que se ha construido un intervalo de confianza a 95% para μ. Entonces, si el valor hipotético de μ según H_0, μ_0, no está en el intervalo de confianza, entonces H_0: $\mu = \mu_0$ se rechaza contra H_1: $\mu \neq \mu_0$ al nivel de 5%. Si μ_0 yace en este intervalo, entonces no se rechaza H_0 al nivel de 5%. Observe cómo cualquier valor de μ_0 se puede probar una vez que se construye el intervalo de confianza y, dado que un intervalo de confianza contiene a más de un valor, existen varias hipótesis nulas que no se rechazarán.

Ejemplo C.8

[Subsidios de capacitación y productividad de los trabajadores]

En el ejemplo de Holzer *et al.* se construyó un intervalo de confianza a 95% para un cambio medio en la tasa de desperdicio μ cuando $[-2.28,-.02]$. Dado que el cero se excluyó de este intervalo, se rechazó H_0: $\mu = 0$ contra H_1: $\mu \neq 0$ al nivel de 5%. Este intervalo de confianza a 95% también significa que no se rechazó H_0: $\mu = -2$ al nivel de 5%. De hecho, existe una sucesión de hipótesis nulas que no se rechazan, dado el intervalo de confianza.

Significancia práctica frente a significancia estadística

En los ejemplos que se han dado hasta ahora, se han producido tres tipos de evidencia concernientes a los parámetros poblacionales: las estimaciones puntuales, los intervalos de confianza y las pruebas de hipótesis. Estas herramientas para aprender acerca de los parámetros poblacionales son igualmente importantes. Existe una tendencia comprensible de los estudiantes a enfocarse en los intervalos de confianza y las pruebas de hipótesis, pues son cosas a las que se puede atribuir niveles de confianza o significancia. Pero en cualquier estudio se deben interpretar las *magnitudes* de las estimaciones puntuales.

El signo y magnitud de $\bar{y}$ determina su **significancia práctica** y nos permite analizar la dirección de un efecto de intervención o política, y si el efecto estimado es "grande" o "pequeño". Por otra parte, la **significancia estadística** de $\bar{y}$ depende de la magnitud de su estadístico *t*. Para probar H_0: $\mu = 0$, el estadístico *t* es simplemente $t = \bar{y}/\mathrm{ee}(\bar{y})$. En otras palabras, la significancia estadística depende de la razón entre $\bar{y}$ y su error estándar. En consecuencia, un estadístico *t* puede ser grande debido a que $\bar{y}$ es grande o $\mathrm{ee}(\bar{y})$ es pequeño. En aplicaciones, es importante discutir tanto la significancia práctica como la estadística, y ser conscientes de que una estimación puede ser significativa estadísticamente sin ser especialmente grande en un sentido práctico. Que una estimación sea importante en la práctica depende del contexto así como del juicio propio, así que no existe un conjunto de reglas para determinar la significancia práctica.

Ejemplo C.9

[Efecto del ancho de la autopista en el tiempo de traslado]

Sea *Y* el cambio en el tiempo de traslado, medido en minutos, para los viajantes que viven en un área metropolitana antes de que una autopista se ampliará a después de su ampliación. Suponga que $Y \sim \mathrm{Normal}(\mu,\sigma^2)$. La hipótesis nula de que la ampliación no reduciría el tiempo promedio de traslado es H_0: $\mu = 0$; la alternativa de que redujo el tiempo promedio de traslado es H_1: $\mu < 0$. Suponga una muestra aleatoria de viajeros del tamaño $n = 900$ se obtiene para determinar la efectividad del proyecto de la autopista. El cambio promedio en el tiempo de traslado se calcula como de $\bar{y} = -3.6$ y la desviación estándar es $s = 32.7$; por tanto, $\text{ee}(\bar{y}) = 32.7/\sqrt{900} = 1.09$. El estadístico t es $t = -3.6/1.09 \approx -3.30$, lo cual es muy significativo en términos estadísticos; el valor-p es de aproximadamente .0005. Así, se concluye que la ampliación de la autopista tendría un efecto estadísticamente significativo en el tiempo de traslado.

Si el resultado de la prueba de hipótesis es la que se reportó en el estudio, se podría caer en un error. Reportar sólo la significancia estadística oculta el hecho de que la reducción estimada en el tiempo promedio de traslado, 3.6 minutos, es mínimo. Para ser francos, se debe reportar la estimación puntual de -3.6, junto con la prueba de significancia.

Encontrar estimaciones puntuales que sean estadísticamente significativas sin ser prácticamente trascendentes puede ocurrir cuando se trabaja con muestras grandes. Para analizar por qué sucede esto, es útil tener la siguiente información.

Consistencia de la prueba. Una **prueba consistente** rechaza H_0 con probabilidad más cerca de uno a medida que el tamaño de la muestra aumenta siempre que H_1 sea verdadera.

Otra forma de decir que una prueba es consistente es que, a medida que el tamaño de la muestra tiende a infinito, la potencia de la prueba se acerca cada vez más a la unidad siempre que H_1 sea verdadera. Todas las pruebas que se cubren en este libro tienen esta propiedad. En caso de la prueba de hipótesis acerca de una media poblacional, la consistencia de la prueba se desprende debido a que la varianza de $\bar{Y}$ converge a cero a medida que el tamaño de la muestra aumenta. El estadístico t para probar H_0: $\mu = 0$ es $T = \bar{Y}/(S/\sqrt{n})$. Dado que $\text{plim}(\bar{Y}) = \mu$ y $\text{plim}(S) = \sigma$, se desprende que si, por ejemplo, $\mu > 0$, entonces T aumenta cada vez más (con una probabilidad alta) cuando $n \to \infty$. En otras palabras, no importa qué tan cercana esté μ de cero, casi se puede estar seguro de rechazar H_0: $\mu = 0$ dado un tamaño de muestra suficientemente grande. Esto no dice nada acerca de si μ es grande en un sentido práctico.

C.7 Comentarios sobre la notación

En el repaso de probabilidad y estadística aquí y en el apéndice B, se ha tenido cuidado de utilizar convenciones estándar para denotar variables aleatorias, estimadores y estadísticos de prueba. Por ejemplo, se ha usado W para indicar un estimador (variable aleatoria) y w para denotar una estimación particular (resultado de la variable aleatoria W). Distinguir entre un estimador y una estimación es importante para comprender varios conceptos en las estimaciones y pruebas de hipótesis. No obstante, hacer esta distinción rápidamente se convierte en una carga en el análisis econométrico debido a que los modelos son más complicados: muchas variables aleatorias y parámetros estarán implicados, y apegarse a las convenciones usuales de la probabilidad y estadística requiere muchos símbolos adicionales.

En el texto principal, se utiliza una convención más simple que se emplea ampliamente en econometría. Si θ es un parámetro poblacional, la notación $\hat{\theta}$ ("theta con gorro") se usará para denotar tanto un estimador como una estimación de θ. La notación es útil en cuanto a que ofrece una forma sencilla de unir un estimador al parámetro poblacional que se supone se va a estimar. Por tanto, si el parámetro poblacional es β, entonces $\hat{\beta}$ denota un estimador o estimación de β; si el parámetro es σ^2, $\hat{\sigma}^2$ es un estimador o estimación de σ^2; y así sucesivamente. En ocasiones se analizarán dos estimadores del mismo parámetro, en cuyo caso se necesitará una notación diferente, como $\tilde{\theta}$ ("theta con gorro").

Aunque desechar las convenciones de probabilidad y estadística para indicar estimadores, variables aleatorias y estadísticos de prueba le impone más responsabilidad, no es muy difícil una vez que se ha comprendido la diferencia entre un estimador y una estimación. Si se están analizando las propiedades *estadísticas* de $\hat{\theta}$ como derivar si es insesgada o no, entonces se estará considerando necesariamente a $\hat{\theta}$ como un estimador. Por otra parte, si se escribe algo como $\hat{\theta} = 1.73$, entonces claramente se está denotando un estimador puntual de una muestra dada de datos. La confusión que puede surgir utilizando $\hat{\theta}$ para denotar ambos debería ser mínima una vez que se ha comprendido bien la probabilidad y la estadística.

RESUMEN

Se han analizado temas de estadística matemática que se basan en el análisis econométrico. La noción de un estimador, que es simplemente una regla para combinar datos para estimar un parámetro poblacional, es fundamental. Se han cubierto varias propiedades de los estimadores. Las propiedades más importantes de la muestra pequeña son el insesgamiento y la eficiencia, esta última depende de comparar varianzas cuando los estimadores son insesgados. Las propiedades de las muestras grandes tienen que ver con la secuencia de estimadores obtenida a medida que el tamaño de la muestra crece, y también son importantes en la econometría. Cualquier estimador útil es consistente. El teorema del límite central implica que, en las muestras grandes, la distribución de muestreo de la mayoría de los estimadores es aproximadamente normal.

La distribución muestral de un estimador se puede utilizar para construir intervalos de confianza. Se ve esto en la estimación de la media de una distribución normal y en el cálculo de intervalos de confianza aproximados en casos que no son normales. La clásica prueba de hipótesis, que requiere especificar una hipótesis nula, una hipótesis alternativa y un nivel de significancia, se realiza al comparar un estadístico de prueba con un valor crítico. Alternativamente, se puede calcular un valor-*p* que permita realizar una prueba a cualquier nivel de significancia.

TÉRMINOS CLAVE

Alternativa de dos colas
Alternativa de una cola
Coeficiente de correlación muestral
Covarianza muestral
Desviación estándar de muestreo
Desviación estándar muestral
Distribución de muestreo
Error cuadrático medio (ECM)
Error estándar
Error tipo I
Error tipo II
Estadístico de prueba
Estadístico *t*
Estimación
Estimador
Estimador consistente
Estimador de intervalo
Estimador de máxima verosimilitud
Estimador de mínimos cuadrados
Estimador insesgado
Estimador insesgado de varianza mínima
Estimador sesgado
Hipótesis alternativa
Hipótesis nula
Inconsistente
Intervalo de confianza
Ley de los grandes números (LGN)
Límite en probabilidad
Método de momentos
Muestra aleatoria
Nivel de significancia
Normalidad asintótica
Población
Potencia de una prueba
Promedio muestral
Prueba de dos colas
Prueba de hipótesis
Prueba de una cola
Región de rechazo
Sesgo
Significancia estadística
Significancia práctica
Teorema del límite central
Valor crítico
Valor-*p*
Varianza de muestreo
Varianza muestral

PROBLEMAS

C.1 Sea Y_1, Y_2, Y_3 y Y_4 variables aleatorias independientes, distribuidas idénticamente, de una población con media μ y varianza σ^2. Sea $\bar{Y} = \frac{1}{4}(Y_1 + Y_2 + Y_3 + Y_4)$ el promedio de estas cuatro variables aleatorias.

i) ¿Cuál es el valor esperado y la varianza de $\bar{Y}$ en términos de μ y σ^2?

ii) Ahora, considere un estimador diferente de μ:

$$W = \frac{1}{8}Y_1 + \frac{1}{8}Y_2 + \frac{1}{4}Y_3 + \frac{1}{2}Y_4.$$

Este es un ejemplo de promedio *ponderado* de Y_i. Muestre que W también es un estimador insesgado de μ. Determine la varianza de W.

iii) Con base en las respuestas en los incisos i) y ii), ¿qué estimador de μ se prefiere, $\bar{Y}$ o W?

C.2 Esta es una versión más general del problema C.1. Sean $Y_1, Y_2, \ldots, Y_n$ sean n variables aleatorias no correlacionadas con media común μ y varianza común σ^2. Sea $\bar{Y}$ el promedio muestral.

i) Defina la clase de *estimadores lineales* de μ mediante

$$W_a = a_1Y_1 + a_2Y_2 + \ldots + a_nY_n,$$

donde a_i son constantes. ¿Qué restricción sobre a_i es necesaria para que W_a sea un estimador insesgado de μ?

ii) Determine la $\text{Var}(W_a)$.

iii) Para cualesquiera números $a_1, a_2, \ldots, a_n$, aplica la siguiente desigualdad: $(a_1 + a_2 + \ldots + a_n)^2/n \le a_1^2 + a_2^2 + \ldots + a_n^2$. Use ésta, junto con las partes i) y ii) para mostrar que $\text{Var}(W_a) \ge \text{Var}(\bar{Y})$ siempre que W_a es insesgada, así que $\bar{Y}$ es el *mejor estimador insesgado lineal*. [*Sugerencia:* ¿en qué se convierte la desigualdad cuando a_i satisface la restricción del inciso i)?].

C.3 Sea $\bar{Y}$ el promedio muestral de una muestra aleatoria con media μ y varianza σ^2. Considere dos estimadores alternativos de μ: $W_1 = [(n-1)/n]\bar{Y}$ y $W_2 = \bar{Y}/2$.

i) Muestre que W_1 y W_2 son estimadores insesgados de μ y encuentre el sesgo. ¿Qué sucede a los sesgos cuando $n \to \infty$? Comente cualquier diferencia importante en el sesgo para los dos estimadores a medida que la muestra aumenta.

ii) Determine los límites en probabilidad de W_1 y W_2. {*Sugerencia:* Use las propiedades PLIM.1 y PLIM.2; para W_1, observe que plim $[(n-1)/n] = 1$.} ¿Qué estimador es consistente?

iii) Determine $\text{Var}(W_1)$ y $\text{Var}(W_2)$.

iv) Argumente que W_1 es un mejor estimador que $\bar{Y}$ si μ está "cercana" a cero. (Considere tanto el sesgo como la varianza).

C.4 Para las variables aleatorias positivas X y Y, suponga que el valor esperado de Y dada X es $\text{E}(Y|X) = \theta X$. El parámetro desconocido θ muestra cómo cambia el valor esperado de Y con X.

i) Defina la variable aleatoria $Z = Y/X$. Muestre que $\text{E}(Z) = \theta$. [*Sugerencia:* use la propiedad CE.2 junto con la ley de las esperanzas iteradas, la propiedad CE.4. En particular, primero muestre que $\text{E}(Z|X) = \theta$ y después utilice CE.4].

ii) Use el inciso i) para demostrar que el estimador $W_1 = n^{-1}\sum_{i=1}^{n}(Y_i/X_i)$ es insesgado para θ, donde $\{(X_i,Y_i): i = 1, 2, \ldots, n\}$ es una muestra aleatoria.

iii) Explique por qué el estimador $W_2 = \bar{Y}/\bar{X}$, donde las barras superiores denotan promedios muestrales, no es lo mismo que W_1. Sin embargo, muéstrese que W_2 también es insesgado para θ.

iv) La siguiente tabla contiene a los datos sobre la producción de maíz para varios condados de Iowa. La USDA predice el número de hectáreas de maíz en cada país con base en fotos satelitales. Los investigadores cuentan el número de "pixeles" de maíz en la foto satelital (comparado con, por ejemplo, el número de pixeles de frijol de soya o de terrenos no cultivados) y los utilizan para predecir el número real de hectáreas. Para desarrollar una ecuación de predicción que empleen los condados en general, la USDA aplicó una encuesta a los granjeros en condados seleccionados para obtener las producciones de maíz en hectáreas. Sea Y_i = la producción de maíz en el condado i y X_i = el número de pixeles de maíz en la fotografía satelital para el condado i. Existen $n = 17$ observaciones para ocho condados. Utilice esta muestra para calcular las estimaciones de θ diseñadas en los incisos ii) y iii). ¿Las estimaciones son similares?

Gráfica	Producción de maíz	Pixeles de maíz
1	165.76	374
2	96.32	209
3	76.08	253
4	185.35	432
5	116.43	367
6	162.08	361
7	152.04	288
8	161.75	369
9	92.88	206
10	149.94	316
11	64.75	145
12	127.07	355
13	133.55	295
14	77.70	223
15	206.39	459
16	108.33	290
17	118.17	307

C.5 Sea Y una variable aleatoria de Bernoulli(θ) con $0 < \theta < 1$. Suponga que interesa estimar la *razón probabilística*, $\gamma = \theta/(1 - \theta)$, que es la probabilidad de éxito sobre la probabilidad de fracaso. Dada una muestra aleatoria $\{Y_1, \ldots, Y_n\}$, se sabe que un estimador insesgado y consistente de θ es $\bar{Y}$, la proporción de éxito en n pruebas. Un estimador natural de γ es $G = \bar{Y}/(1 - \bar{Y})$, la proporción de éxitos sobre la proporción de fracasos en la muestra.

i) ¿Por qué G no es un estimador insesgado de γ?

ii) Use PLIM.2(iii) para mostrar que G es un estimador consistente de γ.

C.6 A usted lo contrata el gobernador para estudiar si un impuesto sobre el licor ha reducido el consumo promedio de licor en su estado. Usted puede obtener, para una muestra de individuos seleccionados al azar, la diferencia en el consumo de licor (en onzas) para los años anteriores y posteriores al impuesto. Para la persona i que forma parte de la muestra aleatoria de la población, Y_i denota el cambio en el consumo de licor. Trate a éstas como una muestra aleatoria de una distribución Normal(μ,σ^2) .

i) La hipótesis nula es que no hubo cambio en el consumo de licor. Exprese esto formalmente en términos de μ.

ii) La alternativa es que hubo un declive en el consumo del licor; enuncie la alternativa en términos de μ.

iii) Ahora, suponga que el tamaño de su muestra es $n = 900$ y que obtiene las estimaciones $\bar{y} = -32.8$ y $s = 466.4$. Calcule el estadístico t para probar H_0 contra H_1; obtenga el valor-p para la prueba. (Debido al tamaño grande de la muestra, sólo use la distribución normal estándar tabulada en la tabla G.1). ¿Rechaza H_0 al nivel de 5%? ¿Al nivel de 1%?

iv) ¿Diría usted que la disminución estimada en el consumo es grande en magnitud? Comente sobre la significancia estadística contra la significancia práctica de esta estimación.

v) ¿Qué se ha supuesto implícitamente en su análisis acerca de otras determinantes del consumo de licor durante el periodo de dos años con el fin de inferir la causalidad del cambio fiscal en el consumo de licor?

C.7 La nueva administración de una pastelería afirma que ahora los trabajadores son más productivos que con la antigua administración, debido a que los salarios hayan "generalmente aumentado". Sea W_i^b el salario del trabajador i bajo la antigua administración y sea W_i^a el salario del trabajador i después del cambio. La diferencia es $D_i \equiv W_i^a - W_i^b$. Suponga que las D_i son muestras aleatorias de una distribución Normal(μ,σ^2).

i) Mediante los siguientes datos sobre 15 trabajadores, construya un intervalo de confianza a 95% para μ.

ii) Formalmente exprese la hipótesis nula de que no ha habido cambio en el promedio de los salarios. En particular, ¿qué es $E(D_i)$ según H_0? Si lo contratan para examinar la validez de la afirmación de la nueva dirección, ¿cuál es la hipótesis alternativa relevante en términos de $\mu = E(D_i)$?

iii) Pruebe la hipótesis nula del inciso ii) frente a la alternativa expresada a los niveles de 5% y 1%.

iv) Obtenga el valor-p para la prueba en el inciso iii).

Trabajador	Salario anterior	Salario posterior
1	8.30	9.25
2	9.40	9.00

(continúa)

Trabajador	Salario anterior	Salario posterior
3	9.00	9.25
4	10.50	10.00
5	11.40	12.00
6	8.75	9.50
7	10.00	10.25
8	9.50	9.50
9	10.80	11.50
10	12.55	13.10
11	12.00	11.50
12	8.65	9.00
13	7.75	7.75
14	11.25	11.50
15	12.65	13.00

C.8 El *New York Times* (2/5/90) reportó las anotaciones de canasta de tres puntos de los mejores diez tiradores en esta categoría de la NBA. La siguiente tabla resume estos datos:

Jugador	IA-AL
Mark Price	429-188
Trent Tucker	833-345
Dale Ellis	1,149-472
Craig Hodges	1,016-396
Danny Ainge	1,051-406
Byron Scott	676-260
Reggie Miller	416-159
Larry Bird	1,206-455
Jon Sundvold	440-166
Brian Taylor	417-157

Nota: *IA* = intentos de anotación y AL anotaciones logradas.

Para un jugador dado, el resultado de un tiro particular se puede representar como una variable Bernoulli (cero-uno): si Y_i es el resultado del tiro i, entonces $Y_i = 1$ si se logró anotar y $Y_i = 0$ si se falló el tiro. Sea θ la probabilidad de hacer cualquier intento de lanzamiento de tres puntos. El estimador natural de θ es $\bar{Y} = FGM/FGA$.

i) Estime θ para Mark Price.

ii) Encuentre la desviación estándar del estimador $\bar{Y}$ en términos de θ y el número de intentos de anotación, n.

iii) La distribución asintótica de $(\bar{Y} - \theta)/\text{ee}(\bar{Y})$ es normal estándar, donde $\text{ee}(\bar{Y}) = \sqrt{\bar{Y}(1 - \bar{Y})/n}$. Use este dato para probar $H_0\colon \theta = .5$ frente a $H_1\colon \theta < .5$ para Mark Price. Use un nivel de significancia a 1%.

C.9 Suponga que un dictador militar en un país determinado realiza un plebiscito (un voto de confianza de sí o no) y afirma que tiene el apoyo de 65% de los votantes. Un grupo de derechos humanos sospecha que hay algún tipo de irregularidad y lo contrata a usted para probar la validez de la afirmación del dictador. Usted tiene un presupuesto que le permite obtener una muestra aleatoria de 200 votantes de ese país.

i) Sea X el número de votos afirmativos obtenidos de una muestra aleatoria de 200 de entre toda la población votante. ¿Cuál es el valor esperado de X si, de hecho, 65% de todos los votantes apoyaron al dictador?

ii) ¿Cuál es la desviación estándar de X, si se supone nuevamente que la fracción verdadera que votó que sí en el plebiscito es .65?

iii) Ahora, usted reúne una muestra de 200, y encuentra que 115 personas en realidad votaron que sí. Use el TLC para aproximar la probabilidad de que encuentre 115 o menos votos de sí de una muestra aleatoria de 200 si, en realidad, 65% de la población entera votó que sí.

iv) ¿Cómo explicaría la relevancia del resultado en el inciso iii) a alguien que no tiene conocimientos de estadística?

C.10 Antes de que una huelga concluyera prematuramente la temporada de la liga mayor de béisbol en 1994, Tony Gwynn de los Padres de San Diego había realizado 165 *hits* en 419 turnos al bate, para un promedio de bateo de .394. Hubo una discusión sobre si Gwynn hubiera podido lograr .400 ese año. Este aspecto puede expresarse en términos de la probabilidad de que Gwynn consiguiera un *hit* en un bateo determinado, llámelo θ. Sea Y_i el indicador de Bernoulli(θ) igual a la unidad si Gwynn hace un *hit* durante su i-ésimo bateo y cero si no. Así, $Y_1, Y_2, \ldots, Y_n$ es una muestra aleatoria de una distribución de Bernoulli(θ) en la que θ es la probabilidad de éxito, y $n = 419$.

La mejor estimación puntual de θ es el promedio de bateo de Gwynn, el cual es sólo la proporción de éxitos: $\bar{y} = .394$. Mediante el hecho de que $\text{ee}(\bar{y}) = \sqrt{\bar{y}(1 - \bar{y})/n}$, constrúyase un intervalo de confianza aproximado a 95% para θ, utilizando la distribución normal estándar. ¿Se diría que ésta es una evidencia contundente en contra de que Gwynn fuera un bateador potencial de .400? Explique.