Unidad 10

- Pruebas de significación.

9

pruebas de significación

Objetivos del capítulo

Después de estudiar este capítulo, se deberá estar en condiciones de:

1. Explicar qué es la prueba de significación y en qué difiere de la estimación.
2. Definir términos tales como hipótesis nula, hipótesis alternativa, nivel de significación, error de tipo I, error de tipo II y valores críticos.
3. Describir cómo las pruebas de significación utilizan las distribuciones de muestreo.
4. Explicar el significado de la frase "división de la distribución de muestreo".

Sinopsis del capítulo

INTRODUCCIÓN

EL OBJETO de este capítulo es *presentar la lógica* de las pruebas de significación sin entrar en detalles con respecto a pruebas específicas. La razón de esto es que los conceptos son muy similares en el caso de todas las pruebas de significación. En los Capítulos 10, 13 y 14, se profundizará en esos detalles. Sin embargo, es fácil dedicarle demasiada atención a los detalles específicos de diversas pruebas, dando lugar a que todos los aspectos de la prueba de significación no queden muy claros.

La prueba de significación y la estimación son dos de las ramas principales de la inferencia estadística. En tanto que el objetivo de la estimación es calcular el valor de cierto parámetro de población, la finalidad de la prueba de significación es decidir si una afirmación acerca de un parámetro de población es verdadera. Por ejemplo, es posible desear determinar si afirmaciones como las siguientes son ciertas:

1. El tiempo promedio de terminación de este examen es de 80 minutos.
2. El 3% de la producción está defectuosa.
3. Se trata de una moneda común (es decir, $P(H) = P(T) = 0.50$).

Ocasionalmente quizá se quiera evaluar una afirmación que no especifique realmente el valor del parámetro en cuestión:

4. El porcentaje del desempleo es el mismo en dos ciudades vecinas.
5. La razón promedio de kilómetros por litro es la misma para los tres tipos de gasolina.

> El *objeto de la prueba de significación* es evaluar proposiciones o afirmaciones acerca de los valores de los parámetros de población.

A partir del análisis sobre la estimación se sabe que los valores estadísticos de la muestra, tales como medias y proporciones, pueden servir como estimaciones de punto de los correspondientes parámetros de población. Asimismo se observó que, debido a la variabilidad del muestreo, asociada al muestreo al azar, los valores estadísticos de la

muestra tienden a *aproximar* en lugar de a *igualar* los parámetros poblacionales. Cabe tener esto en cuenta para el análisis de una proposición o afirmación respecto al valor de un parámetro de una población que emplee datos muestrales.

De ahí que el aspecto principal de la prueba de significación sea determinar si la diferencia entre un valor propuesto de un parámetro de población y el valor estadístico de la muestra se debe razonablemente a la variabilidad del muestreo, o si la discrepancia es demasiado grande para ser considerada de esa manera.

Es probable que para las pruebas de significación se pueda apreciar mejor el método básico si se plantea un problema sencillo. Considérese la siguiente situación: Se inspecciona una muestra de 142 productos de una enorme remesa y se observa que el 8% de ellos está defectuoso. El proveedor de dichos productos garantizó que un porcentaje no mayor al 6% de cualquier cargamento tendría defectos. La pregunta que se habrá de contestar mediante la prueba de significación es si la afirmación pronunciada por el proveedor es verdadera.

El primer paso de la prueba de significación es formular dos hipótesis con respecto a dicho aserto. Las hipótesis son explicaciones potenciales (teorías) que intentan informar acerca de hechos observados en situaciones en las que existen algunos factores desconocidos. En este caso, la incógnita es el porcentaje verdadero de productos defectuosos que hay en el embarque. El hecho conocido es que una muestra aleatoria señaló un 8% defectuoso. Una hipótesis posible sería que el porcentaje real de partes defectuosas del lote sea mayor del 6% propuesto. Otra sería que la proposición es verdadera. Ahora, si ésta es cierta, ¿cuál sería la causa del hecho de que una muestra señalara un 8% de partes defectuosas? Una posibilidad es que la causa sea la variabilidad del muestreo.

Ahora se definirán más formalmente los dos tipos de hipótesis que se requiere formular. La que señala que la proposición es verdadera recibe el nombre de *hipótesis nula* y se representa mediante el símbolo H_0; y la segunda, que afirma que la proposición es falsa se denomina *hipótesis alternativa* y se designa mediante el signo H_1.

> La *hipótesis nula*, H_0, es un enunciado que expresa que el parámetro de la población es como se especificó (es decir, que la proposición es verdadera).
> La *hipótesis alternativa*, H_1, es un enunciado que ofrece una alternativa a la proposición (por ejemplo, el parámetro es mayor que el valor propuesto).

En el ejemplo anterior, la hipótesis nula es la que establece que el porcentaje verdadero de productos defectuosos es el 6%, lo cual se representaría mediante:

$$H_0: p = 6\%$$

La alternativa es que el porcentaje de productos defectuosos p es mayor del 6%, lo cual se enunciaría como:

$$H_1: p > 6\%$$

Si la decisión después de efectuar el análisis es aceptar H_0, esto significa que la discrepancia entre el porcentaje de productos defectuosos observado en la muestra y el porcentaje de elementos defectuosos de la población (propuesto) se debe probablemente a la variación casual del muestreo. Por el contrario, la decisión de rechazar H_0 significaría que la variación entre el valor observado y el propuesto es demasiado grande como para deberse únicamente a la casualidad.

¿VARIACION CASUAL O VARIACION REAL?

La "prueba" consiste en saber si algún valor estadístico muestral observado puede provenir razonablemente de una población que presente el parámetro propuesto. De ahí que se quiera considerar la variabilidad de muestreo que pudiera surgir, dada una población como la establecida.

El segundo paso en la prueba de significación es identificar la distribución de muestreo adecuada, ya que ésta describirá ampliamente la variación casual. En este ejemplo, en el cual se trabaja con proporciones muestrales y una muestra grande ($n = 142$), la distribución de muestreo apropiada es una distribución normal que tiene una media de p y una desviación estándar de

$$\sigma_p = \sqrt{\frac{p(1-p)}{n}}$$

en la cual,

p = proporción de la población
n = tamaño de la muestra

De este modo, *si* la afirmación es verdadera, la proporción de la muestra del 8% proviene de una distribución de muestreo que tiene una media del 6% y una desviación estándar de

$$\sigma_p = \sqrt{\frac{0.06(0.94)}{142}} = 0.2$$

Ahora se puede observar que la discrepancia del 2% radica en una desviación estándar que está por encima del valor esperado, suponiendo que 0.06 es la proporción verdadera de la población:

$$z = \frac{0.08 - 0.06}{0.02} = +1.0$$

Además, la probabilidad de obtener una discrepancia *mayor* del 8% con una muestra de 142 observaciones obtenidas a partir de una población con una proporción del 6% es aproximadamente del 16%, según se indica en la figura 9.1. Esto parece sugerir que el factor casual por sí solo *podría* dar lugar a discrepancias. Es innecesario decir que *no se puede* establecer en forma definitiva que la población muestreada presente el 6% de elementos defectuosos, pero, en vista de la distribución de muestreo en una población de esta naturaleza y del valor estadístico observado, la proposición sí parece razonable.

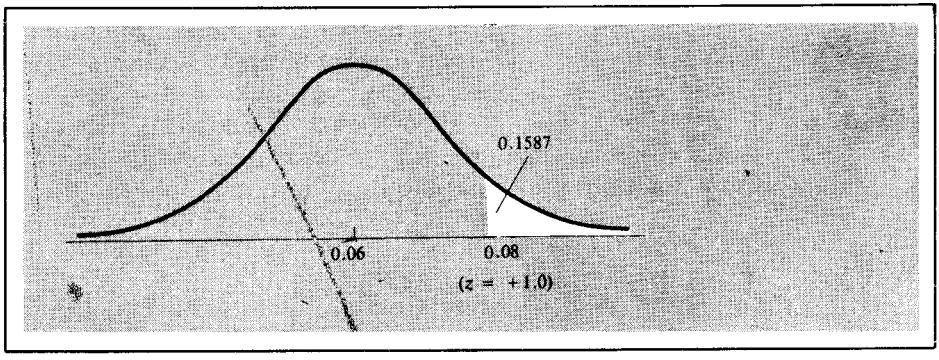

FIGURA 9.1 La probabilidad de una proporción muestral superior al 8%, dada una proporción de población del 6%, es 0.1587

Por otra parte, si se obtuvo una proporción muestral de, digamos 19%, entonces

$$z = \frac{0.19 - 0.06}{0.02} = +6.5$$

y parecería sumamente improbable que este valor estadístico muestral pudiera provenir de una población con un parámetro propuesto de 6%. En ese caso se estaría más inclinado a rechazar H_0. En la figura 9.2 se ilustra una comparación entre las dos posibilidades.

Sin embargo, no todas las situaciones serán tan evidentes como para poder "visualizarlas" de esta manera. Por tanto, se requerirá de un método más riguroso para resolver el problema. La pregunta es: ¿dónde se trazará la línea divisoria entre lo que se puede considerar razonablemente como "variación casual" y lo que se puede tomar como "variación significativa"?

Al intentar dar respuesta a esta pregunta, se debe considerar lo siguiente: Aproximadamente el 5% de los valores estadísticos de la muestra en una distribución normal da-

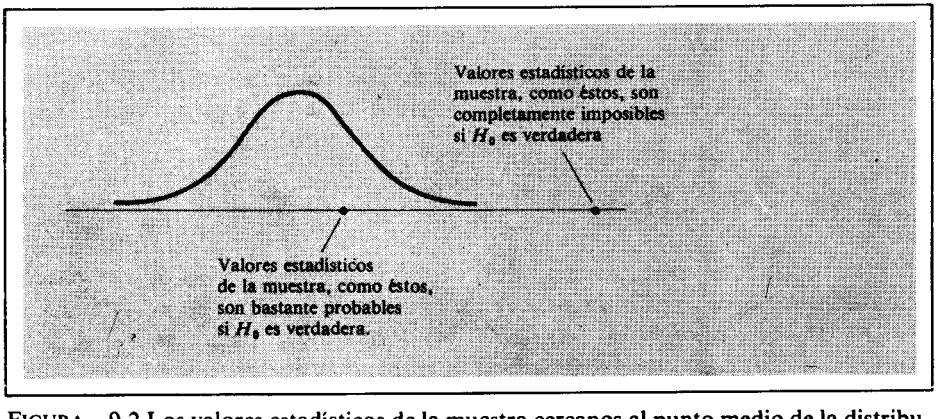

FIGURA 9.2 Los valores estadísticos de la muestra cercanos al punto medio de la distribución de muestreo tienden a apoyar esta proposición, en tanto que los más lejanos a este punto indican que es falsa.

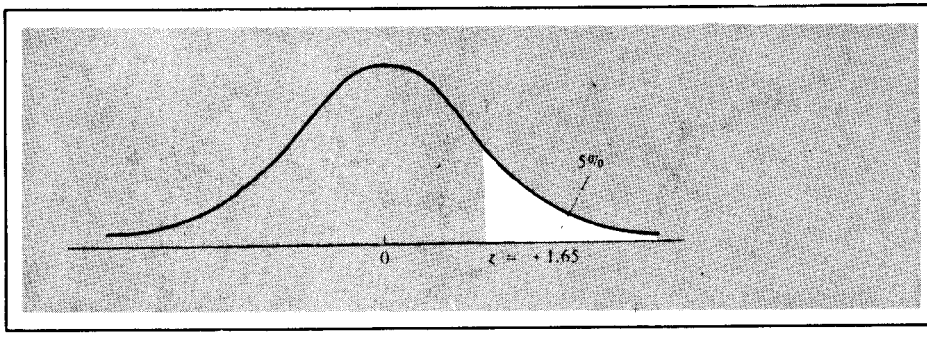

FIGURA 9.3 Aproximadamente el 5% de los valores estadísticos de la muestra producirá un valor de z superior cuando la hipótesis nula es verdadera.

rá como resultado un valor de z mayor que $+1.65$ (ver figura 9.3). Así, aunque el valor de la muestra *esperado* es de 6%, el 5% de los posibles valores estadísticos muestrales será superior a $p + 1.65\sigma_p$. Por tanto, si establecemos que $z = +1.65$ es la línea divisoria, existe un 5% de riesgo de rechazar H_0 cuando este valor resulte realmente verdadero. Otra posibilidad sería utilizar $z = +2.33$ como el "valor crítico", dado que únicamente habría una posibilidad de alrededor del 1% de observar un valor estadístico de muestra más extremo que cuando H_0 es verdadera. Por supuesto, la distribución de muestreo teórico se extiende más allá del infinito, de manera que se debe trazar la línea en alguna parte. Además, se está de acuerdo en que ciertos valores podrían parecer tan improbables que se les rechazaría inmediatamente. Los valores críticos que generalmente se seleccionan en las pruebas de significación son los que comprenden riesgos del 5%, 2.5% o 1% de rechazar H_0 cuando sea verdadera. La probabilidad de rechazar una hipótesis nula verdadera recibe el nombre de nivel de significación de una prueba, y se le designa mediante el símbolo α (alfa griega).

> El *nivel de significación* de una prueba es la probabilidad de rechazar una hipótesis nula que sea verdadera.

Por tanto, el tercer paso en una prueba de este tipo es seleccionar un nivel de significación que sea aceptable. Esto, a su vez, indicará un valor crítico correspondiente que servirá como un estándar de comparación respecto al cual juzgar un "valor estadístico de prueba" observado (es decir, la proporción de la muestra del 8% tiene un valor de z_{prueba} de $+1.0$).

Por tanto, la esencia de una prueba de significación es dividir una distribución de muestreo con base en el supuesto de que H_0 es verdadera, en regiones de aceptación y rechazo respecto de H_0. Un valor crítico se selecciona con base en una probabilidad especificada de que el tomador de decisiones esté dispuesto a aceptar o rechazar una H_0 verdadera. Un valor estadístico de prueba se calcula a partir de datos de la muestra y

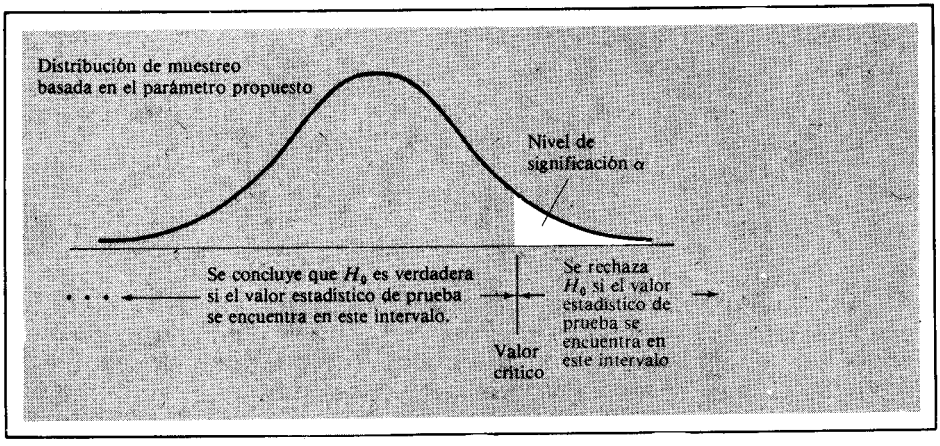

FIGURA 9.4 La distribución de muestreo está dividida en regiones de aceptación o de rechazo, teniendo al valor crítico como punto divisor.

del valor esperado (propuesto), el cual es comparado con el valor crítico. Un valor estadístico de prueba superior al valor crítico señala que se debe rechazar H_0 (es decir, que la sola variabilidad del muestreo no tendrá en cuenta el valor estadístico de muestra observado), mientras que un valor de prueba menor que el valor crítico indica que se debe aceptar H_0. El concepto total se ilustra en la figura 9.4.

PRUEBAS UNI Y BILATERAL

El interés por detectar desviaciones no aleatorias, (es decir, significativas) a partir de un parámetro especificado, puede comprender desviaciones en ambas direcciones o en una sola. De ese modo, al lanzar una moneda se puede considerar que está arreglada si caen demasiadas o muy pocas caras. La hipótesis alternativa sería simplemente que la moneda está arreglada, por lo que se tendrían que investigar las desviaciones en ambas direcciones. Pero si se apostara en favor de las caras, el interés radicaría solamente en no obtener demasiadas pocas caras. Por ello, la hipótesis alternativa sería que cayeran muy pocas caras (es decir, que la probabilidad de caras fuera menor que 0.50), y al efectuar una evaluación nos concentraríamos solamente en el tipo de desviación no aleatoria proveniente del número esperado de caras.

En esencia, la hipótesis alternativa se utiliza para indicar qué aspecto de variación no aleatoria resulta de interés. Existen tres casos posibles: 1) concentrarse en *ambas direcciones;* 2) concentrarse en desviaciones por *debajo* del valor esperado; o, 3) concentrarse en desviaciones *por encima* del valor esperado. Simbólicamente, en lo referente al ejemplo del lanzamiento de una moneda, estos tres casos se pueden representar de la siguiente manera:

$$H_0: p = 0.50$$

Caso 1. $H_1: p \neq 0.50$ (ambas direcciones: demasiadas o muy pocas).
Caso 2. $H_1: p < 0.50$ (desviación por abajo: muy pocas caras).
Caso 3. $H_1: p > 0.50$ (desviación por arriba: demasiadas caras).

"¿Llamarías significativo a esto?"

Obsérvese que la hipótesis nula se representa de la misma forma, independientemente de cuál sea la hipótesis alternativa.* La diferencia entre estos casos se ilustra en la figura 9.5. Cabe observar que para una prueba de dos colas, el área de cada una es $\alpha/2$. En el primer caso, un valor demasiado por encima o demasiado por debajo del valor esperado causaría un rechazo de la hipótesis nula. Sin embargo, en el segundo caso, únicamente un valor demasiado abajo del valor esperado rechazaría la hipótesis nula. En el tercer caso, ocurre justamente lo opuesto, dado que sólo valores muy por encima del valor esperado causan rechazo.

En la práctica, se utiliza la prueba bilateral siempre que la divergencia en *ambas* direcciones sea crítica, como podría ser en el caso de la fabricación de ropa, donde las camisas que sean demasiado grandes o demasiado pequeñas no corresponderán con una talla establecida. Otro ejemplo es el caso donde hay dos partes que deben embonar o ajustar perfectamente, como cuando se trata de un tornillo y su tuerca. Una variación excesiva ocasionará un ajuste tan holgado que impida que permanezcan unidas las piezas, o tan apretado que no puedan embonar de ningún modo.

*En realidad desde el punto de vista de la compleción, en el caso de una prueba de cola, la hipótesis nula se deberá aplicar a cierto intervalo de valores. Por ejemplo, si H_1 es $p<0.5$, entonces H_0 será $p \geq 0.5$, y para H_1: $p>0.5$ H_0 será $p \leq 0.5$. Sin embargo, la base para la distribución de muestreo que se utiliza para la prueba no puede ser un intervalo de valores, sino que tiene que ser un *sólo* valor. Por esa razón, en este libro se tratará de especificar un valor único para H_0.

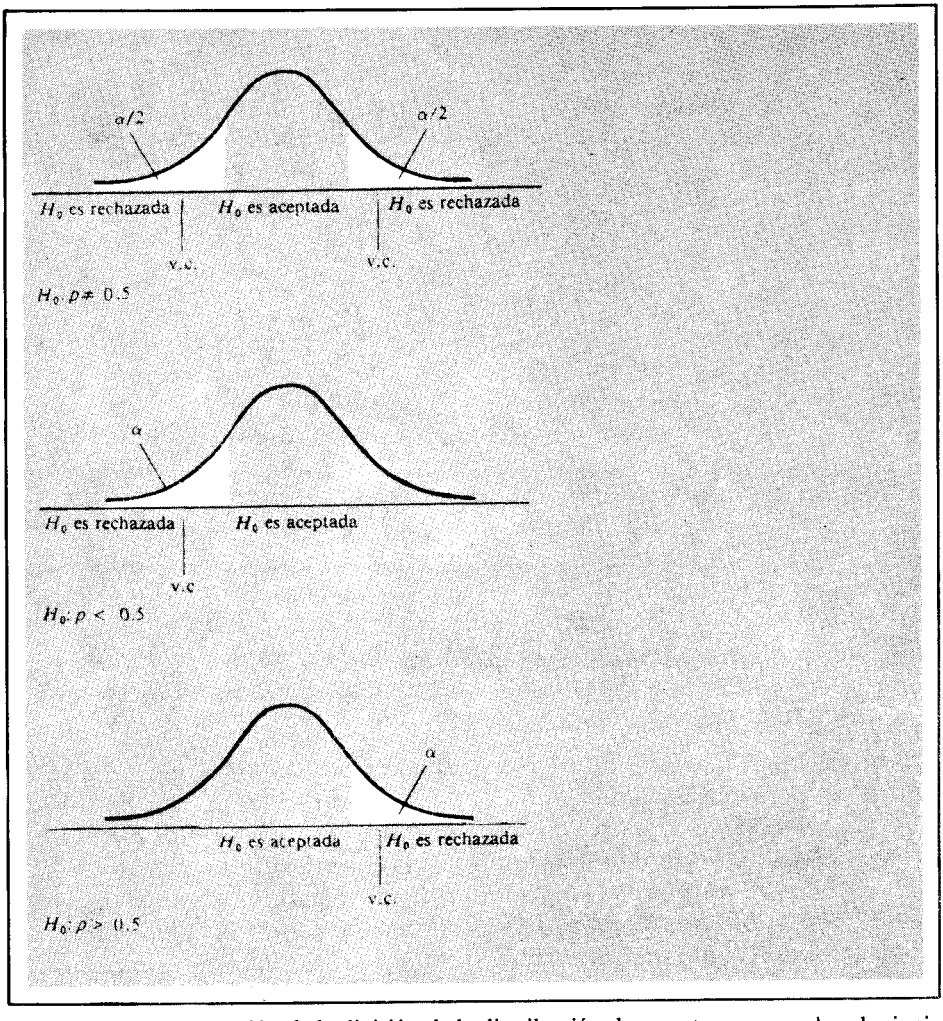

FIGURA 9.5 Comparación de la división de la distribución de muestreo en pruebas de significación unilaterales o bilaterales. En las pruebas unilaterales obsérvese que los signos $>$ o $<$ apuntan hacia la cola que está siendo utilizada.

La prueba de cola izquierda es útil cuando se quiere observar si se ha cumplido un estándar *mínimo*. Algunos ejemplos de ello son: Un mínimo de grasa en la leche entera, el peso neto de productos empacados, la resistencia a la tensión por parte de los cinturones de seguridad y la vida del producto, según lo especificado en la garantía. Una prueba de cola derecha sirve cuando estándares *máximos* no deben ser excedidos. Algunos ejemplos son: la cantidad de grasa permitida en la leche descremada, la radiación emitida por estaciones o plantas nucleares, el número de artículos defectuosos en un embarque de productos manufacturados y el grado de contaminación del aire producido por una chimenea.

ERRORES DE TIPO I Y DE TIPO II

Existen dos tipos de errores que son inherentes al proceso de la prueba de significación. Ya se ha observado que el creer que H_0 es falsa cuando realmente es verdadera, puede conllevar cierto riesgo. La probabilidad de cometer este error es igual al nivel de significación de una prueba, α. También se conoce como *error de tipo I*. Un segundo tipo de error que también se puede presentar, es aceptar H_0 cuando no es verdadera. Este recibe el nombre de *error de tipo II* y se designa mediante el símbolo β (beta griega).

Naturalmente existe la esperanza de que H_0 sea aceptada cuando sea verdadera, y rechazada cuando sea falsa. Por tanto, en cualquier prueba pueden presentarse cuatro posibilidades. En la tabla 9.1 se establece una comparación entre ellas. Es importante saber que una vez que se toma una decisión, ésta puede ser correcta o incurrir en *algún* tipo de error, y la decisión (aceptada o rechazada) indicará qué tipo de error es posible. Obsérvese también que, cuando H_0 es verdadera no puede haber un error de tipo II, y que cuando H_0 es falsa no se puede cometer un error de tipo I.

Se comete un error de tipo I si se rechaza H_0 cuando es verdadera. La probabilidad de un error de tipo I es igual al nivel de significación de una prueba de hipótesis.

Se comete un error de tipo II si se acepta H_0 cuando no es verdadera.

TABLA 9.1 Errores de tipo I y de tipo II

		Si H_0 es	
		Verdadera	Falsa
Y se toma esta acción	H_0 es aceptada	Decisión correcta	Error de tipo II β
	H_0 es rechazada	Error de tipo I α	Decisión correcta

Anteriormente se mencionó que se podría reducir la probabilidad de rechazar erróneamente H_0, seleccionando valores críticos extremos [es decir, que deje poca área en (el) los extremo(s) de la distribución]. Sin embargo, existe una relación inversa entre los errores de tipo I y de tipo II: si disminuye la probabilidad de un error de tipo I aumen-

tará la probabilidad de incurrir en uno de tipo II.* En forma ideal, se deberá reducir al mínimo el *costo* de cometer un error de tipo I, respecto al de incurrir en uno de tipo II, aunque en la práctica es común seleccionar tradicionalmente niveles de errores de tipo I y pasar por alto los de tipo II.

RESUMEN

En este capítulo se analizó el concepto general de una prueba de significación sin profundizar en los detalles de la misma. En los capítulos siguientes se presentarán diversos tipos de dichas pruebas. Sin embargo, en su mayor parte, los conceptos subyacentes serán esencialmente los mismos que los que se estudian en este capítulo.

Las pruebas de significación se utilizan para evaluar las proposiciones acerca de parámetros de población. El procedimiento general es el siguiente:

1. Formular las hipótesis nula y alternativa.
2. Seleccionar la distribución de muestreo apropiada.
3. Seleccionar un nivel de significación (y de ese modo, valores críticos).
4. Calcular un valor estadístico de prueba y compararlo con los valores críticos.
5. Rechazar la hipótesis nula si el valor estadístico de la prueba excede los valores críticos; aceptarla si se presenta el caso contrario.

El aspecto central de todo el proceso es una distribución de muestreo basada en la premisa de que la proposición es verdadera. Esto indica el grado en el que pueden variar los resultados muestrales, debido simplemente a la variación casual en el muestreo. La distribución de muestreo se divide en una región que sugiere que se debe aceptar H_0 y una (prueba de una cola) o dos (prueba de dos colas) regiones donde se debe rechazar H_0. La probabilidad de cometer el error de tipo I recibe el nombre de nivel de significación de una prueba, y es igual al área de la región rechazada. Se comete un error de tipo II si se acepta H_0 cuando es falsa.

En este capítulo se utilizó una prueba que requiere una distribución de muestreo normal. En capítulos posteriores se presentarán pruebas que utilizan otras distribuciones. La elección de la distribución de muestreo depende de cosas tales como el tipo de los datos por analizar (medidas, rangos o categorías), tamaño de la muestra, y si se pueden establecer ciertos supuestos con respecto a la población subyacente (por ejemplo, ¿es normal la población?). Los siguientes capítulos están organizados bajo la consideración principal del tipo de datos que se han analizado. Por ejemplo, en el siguiente capítulo se trabaja con pruebas de medias, y en estas pruebas se utiliza la medición de datos.

TEMAS DE REPASO

1. Defina brevemente cada uno de los conceptos siguientes:
   a. hipótesis
   b. nivel de significación

* Determinar la probabilidad de un error de tipo II es menos sencillo que calcular la de un error de tipo I. Este tema se comentará al final del siguiente capítulo.