Unidad 12

- Pruebas de significación para proporciones.

12

pruebas de significación para proporciones

Objetivos del capítulo

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

1. Explicar el objeto de las pruebas de proporciones.
2. Explicar la finalidad de las pruebas de proporciones de una muestra, dos muestras y de k muestras.
3. Resolver los problemas comunes, utilizando las técnicas de este capítulo
4. Comentar brevemente la distribución ji o chi cuadrada y compararla con las distribuciones que se estudiaron anteriormente.
5. Utilizar la tabla de ji cuadrada para obtener probabilidades.

Sinopsis del capítulo

INTRODUCCIÓN

LAS PRUEBAS DE proporciones son adecuadas cuando los datos que se están analizando constan de cuentas o frecuencias de elementos de dos o más clases. El objeto de estas pruebas es evaluar las afirmaciones con respecto a una proporción (o porcentaje) de población. Las pruebas se basan en la premisa de que una proporción muestral (es decir, x ocurrencias en n observaciones, o x/n) será igual a la proporción verdadera de la población si se toman márgenes o tolerancias para la variabilidad muestral. Las pruebas suelen enfocarse en la diferencia entre un número esperado de ocurrencias, suponiendo que una afirmación es verdadera, y el número observado realmente. La diferencia se compara con la variabilidad prescrita mediante una distribución de muestreo que tiene como base el supuesto de que H_0 es realmente verdadera.

En muchos aspectos, las pruebas de proporciones se parecen a las pruebas de medias, excepto que, en el caso de las primeras, los datos muestrales se consideran como cuentas en lugar de como mediciones. Por ejemplo, las pruebas para medias y proporciones se pueden utilizar para evaluar afirmaciones con respecto a 1) un parámetro de población único (prueba de una muestra), 2) la igualdad de parámetros de dos poblaciones (prueba de dos muestras) y 3) la igualdad de parámetros de más de dos poblaciones (prueba de k muestras). Además, para tamaños grandes de muestras, la distribución de muestreo adecuada para pruebas de proporciones de una y dos muestras es aproximadamente normal, justo como sucede en el caso de pruebas de medias de una y dos muestras.

La presentación de pruebas de significación para proporciones de muestras se divide en dos partes. La primera nos señala las pruebas de una y dos muestras. La segunda parte nos presenta pruebas de más de dos proporciones, y utiliza una distribución de muestreo que aún no ha sido estudiada aquí.

PRUEBA DE PROPORCIONES DE UNA MUESTRA

Cuando el objeto del muestreo es evaluar la validez de una afirmación con respecto a la proporción de una población, es adecuado utilizar una prueba de una muestra. La metodología de prueba depende de si el número de observaciones de la muestra es grande o pequeño. Con muestras de más de 20 observaciones, la distribución normal es aceptable; con tamaños más pequeños de muestras se deberá utilizar la distribución binomial. En primer lugar se considerará el caso de la muestra grande y después el de la muestra pequeña.

Como se habrá observado anteriormente, las pruebas de grandes muestras de medias y proporciones son bastante semejantes. De este modo, los valores estadísticos de prueba para ambos tipos miden la desviación de un valor estadístico de muestra a partir de un valor propuesto. Y ambas pruebas se basan en la distribución normal estándar para valores críticos. Quizá la única diferencia real entre las ambas radica en la forma como se obtiene la desviación estándar de la distribución de muestreo.

Cabe recordar que no hay relación entre la media y la desviación estándar de una distribución de muestreo de medias, pero que no sucede lo mismo en el caso de la media (p) y la desviación estándar ($\sigma_p = \sqrt{p(1-p)/n}$), de una distribución de muestreo de proporciones. Sin embargo, a diferencia del método que se utiliza en la estimación, en el cual la proporción de la *muestra* fue incluida en la fórmula, el valor de p utilizado para calcular la desviación estándar se basó en el valor establecido en la H_0.

Por ejemplo, la hipótesis nula puede ser

$$H_0: p_0 = 0.20$$

Entonces, para calcular σ_{p0}, se podría utilizar el valor 0.20, así como el tamaño de muestra n. De este modo, supóngase que $n = 100$. Por tanto

$$\sigma_{p_0} = \sqrt{\frac{(0.2)(1-0.2)}{100}} = 0.04$$

El símbolo p_0 se utiliza para denotar el valor especificado en H_0. La prueba comprende el cálculo del valor estadístico de prueba z:

$$z = \frac{\text{proporción de la muestra} - \text{proporción propuesta}}{\text{desviación estándar de la proporción}} = \frac{(x/n) - p_0}{\sqrt{p_0(1-p_0)/n}}$$

Posteriormente este valor es comparado con el valor de z, obtenido a partir de una tabla normal a un nivel de significación seleccionado.

Como ocurrió con la prueba de medias de una muestra, las pruebas de proporciones pueden ser de una o dos colas. El tipo de prueba refleja H_1. Por ejemplo, hay tres posibilidades para H_1:

$$H_1: p > p_0 \qquad H_1: p < p_0 \qquad H_1: p \neq p_0$$

La primera alternativa establece una prueba de cola derecha, la segunda, una de cola izquierda y la tercera, una prueba de dos colas. Los tres ejemplos siguientes demuestran estas pruebas.

Ejemplo 1 Un fabricante asegura que un embarque de ciertos clavos contiene menos

del 1% de piezas defectuosas. Una muestra aleatoria de 200 de ellos presenta 4 defectuosos (es decir, el 2%). Pruebe esta afirmación al nivel 0.01.

Solución:

La hipótesis nula es

$$H_0: p_0 = 1\%$$

La hipótesis alternativa podría ser razonablemente

$$H_1: p_0 > 1\%$$

dado que se desea evitar que se acepte una remesa con más del 1% de clavos defectuosos, sin importar si la remesa está mejor de lo esperado. Por tanto, se utilizará una prueba de una cola, cuya línea divisoria está entre aceptar H_0 y rechazar H_0 en la cola derecha.

La desviación estándar de la distribución de muestreo si H_0 es realmente verdadera, es

$$\sigma_{p_0} = \sqrt{\frac{p_0(1 - p_0)}{n}} = \sqrt{\frac{0.01(0.99)}{100}} = 0.007$$

Así, el valor estadístico de prueba se puede calcular:

$$z = \frac{0.02 - 0.01}{0.007} \approx +1.43$$

El valor tabular normal, utilizando el nivel 0.01 y la tabla normal, es $z = +2.33$. En consecuencia, H_0 es aceptada, como se ilustra en la figura 12.1.

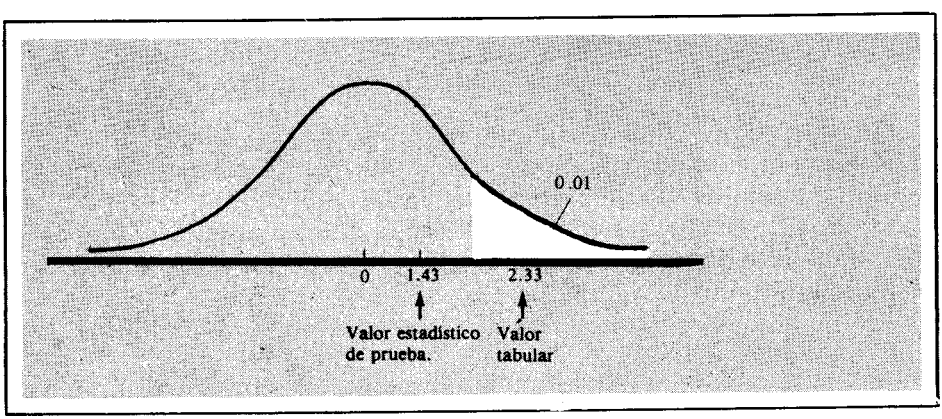

FIGURA 12.1 La evaluación del valor estadístico de prueba da lugar a la aceptación de H_0.

Ejemplo 2 En un estudio se afirma que 9 de 10 médicos (es decir, el 90% recomiendan la aspirina a pacientes que tienen hijos. Pruebe esta aseveración, a un nivel de signifi-

cación 0.05, respecto a la alternativa de que la proporción real de los médicos que hacen esto es menor del 90%, si una muestra aleatoria de 100 médicos revela que 80 de ellos recomiendan la aspirina.

Solución:

$$H_0: p_0 = 90\%$$
$$H_1: p_0 < 90\%$$

La desviación estándar de la distribución de muestreo, suponiendo que H_0 es verdadera, es

$$\sigma_{p_0} = \sqrt{\frac{0.90(0.10)}{100}} = 0.03$$

El valor estadístico de prueba es

$$z = \frac{0.80 - 0.90}{0.03} = \frac{-0.10}{0.03} 3 -3.33$$

El valor tabular de la curva normal (nivel 0.05) es $z = -1.65$. Por tanto, se rechaza H_0 y se concluye que *menos del* 90% de los médicos recomiendan la aspirina. Ver la figura 12.2.

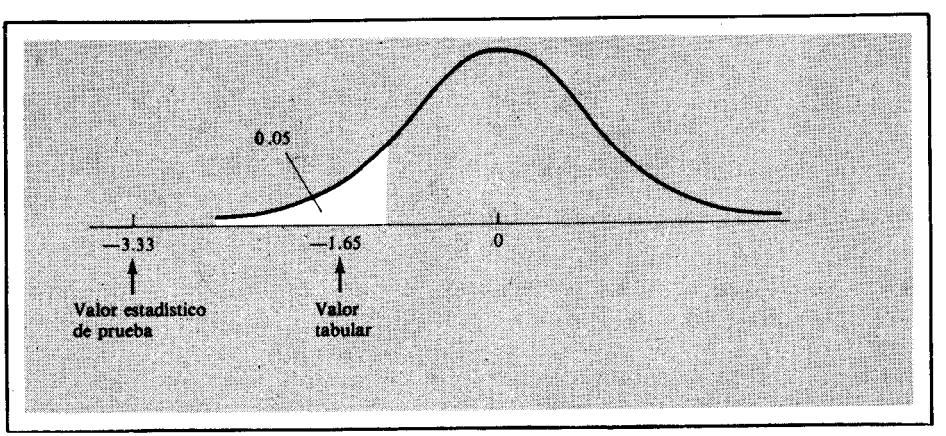

FIGURA 12.2 La hipótesis nula es rechazada

Ejemplo 3 En un periódico se afirma que aproximadamente el 25% de los adultos en su área de circulación son analfabetos según las evaluaciones del gobierno. Ponga a prueba esta aseveración contra la alternativa de que el porcentaje verdadero no es el 25%, y utilice una probabilidad del 5% de que se cometa un error de tipo I. Una muestra de 740 personas indica que el 20% se considerarían analfabetas utilizando las mismas normas.

Solución:

$$H_0: p_0 = 25\%$$
$$H_1: p_0 \neq 25\%$$

La desviación estándar de la distribución de muestreo es

$$\sigma_{p_0} = \sqrt{\frac{p_0(1-p_0)}{n}} = \sqrt{\frac{0.25(0.75)}{740}} \approx 0.016$$

El valor estadístico de prueba es

$$z = \frac{0.20 - 0.25}{0.016} = \frac{-0.05}{0.016} = -3.1$$

El valor tabular normal para 0.05 y una prueba de dos colas es $\pm$ 1.96. En consecuencia, se rechaza H_0 y se concluye que el porcentaje real es menor del 25% como se indica en la figura 12.3.

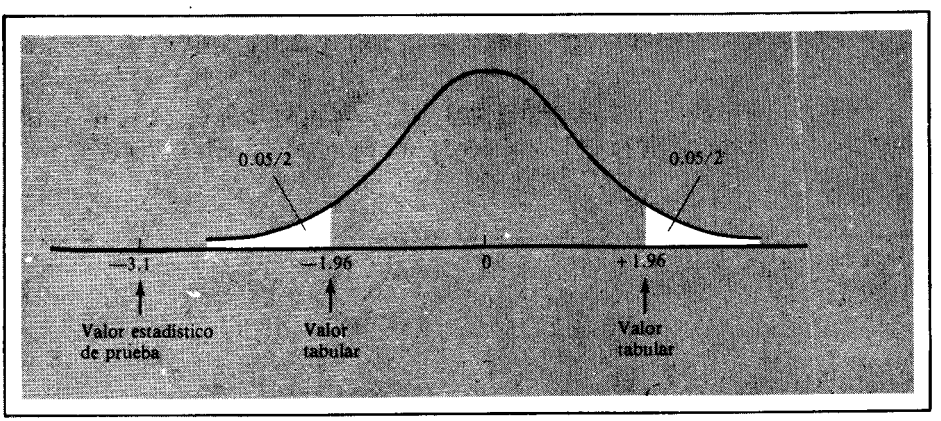

FIGURA 12.3 **La hipótesis nula es rechazada.**

Si se muestrea a partir de una población finita (y $n/N > 5\%$), se debe utilizar el factor finito de corrección $\sqrt{(N-n)/(N-1)}$.

Ejemplo 4 Un importante fabricante de dulces asegura que menos del 3% de las bolsas de lunetas de chocolate están por debajo del nivel de llenado. Una comprobación aleatoria revela que 4 de 50 bolsas se encuentran en esta situación. La muestra fue tomada de una remesa de 400 bolsas de lunetas. ¿Refuta la evidencia muestral la afirmación del fabricante (es decir, hay más del 3% de bolsas que no están completamente llenas)?

Solución:

Es necesario obtener una prueba de cola derecha, ya que se quiere evitar que haya *demasiadas* bolsas que no estén completamente llenas. Como no se proporciona el dato

del nivel de significación, supóngase que es 0.05. La desviación estándar de la distribución de muestreo es

$$\sigma_p = \sqrt{\frac{p_0(1-p_0)}{n}}\sqrt{\frac{N-n}{N-1}} = \sqrt{\frac{0.03(0.97)}{50}}\sqrt{\frac{350}{399}}$$

El valor estadístico de prueba es

$$z = \frac{(x/n) - p_0}{\sigma_p} = \frac{\frac{4}{50} - 0.03}{0.023} = \frac{0.08 - 0.03}{0.023} = 2.17$$

H_0 es rechazada, como se indica en la figura 12.4.

Cuando el número de observaciones es de 20 o menos, se puede utilizar la tabla binomial acumulativa, para evaluar datos de la muestra. Considérese el siguiente ejemplo. El propietario de un cierto establecimiento quiere saber si sus clientes asiduos pueden distinguir entre la cerveza de barril o embotellada. Por tanto, elige 11 clientes y les proporciona a cada uno dos tarros no marcados, uno para cada tipo de cerveza. Le pide a cada cliente que elija el tarro que prefiera. Considera que se mostrará preferencia por la cerveza de barril. Si no hubiera preferencia, el resultado esperado sería que cerca de la mitad elegiría la cerveza de barril y la otra mitad, la cerveza embotellada. Si su hipótesis es correcta, pocos seleccionarán este último tipo de cerveza. Las hipótesis serían entonces:

H_0: ninguna preferencia (es decir, aproximadamente la mitad seleccionaría uno de los tipos de cada cerveza), o bien $np = 11\,(\frac{1}{2})$

H_1: pocos seleccionarán la cerveza embotellada, o bien $np < 11\,(\frac{1}{2})$

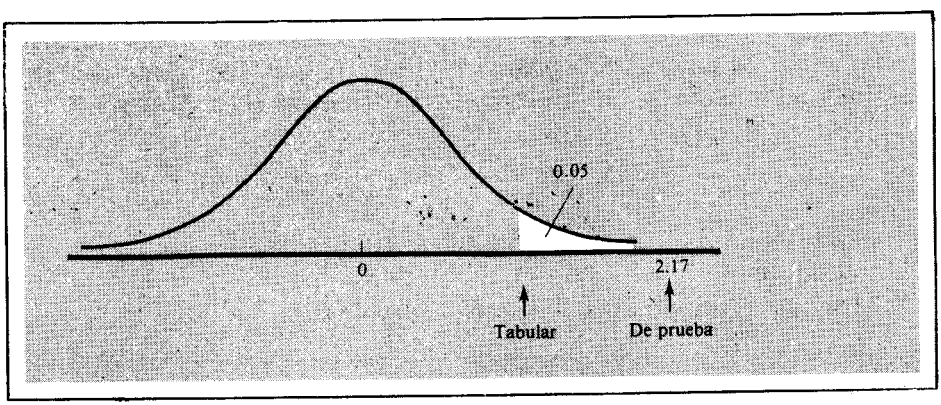

FIGURA 12.4 H_0 es rechazada.

Supóngase que cuatro clientes eligen la cerveza embotellada, y que el resto escoge la de barril. El propietario quiere hacer una prueba a un nivel de 0.05. ¿Apoyan las pruebas a H_0 o a H_1?

A diferencia de las pruebas comentadas anteriormente, no se utilizará un valor crítico. En lugar de ello, se trabajará con probabilidades en las regiones de colas. El motivo de esta desviación proviene del hecho de que las distribuciones discretas tienden a "aglomerarse". Se tendría que elegir un valor crítico, como 1, 2 ó 3, valores que generalmente no proporcionan el nivel de significación deseado. Por ejemplo, cabe observar, a partir de la porción de la tabla binomial acumulativa presentada aquí (con $n = 11$ y $p = 0.50$) que $p\,(x < 2) = 0.0327$. En consecuencia, un valor crítico de "2 o menos selecciones de cerveza embotellada" produciría un P(error de tipo I) más pequeño que el especificado. Por otra parte, un valor crítico de 3 dejaría 0.1133 en la cola, lo cual es demasiado grande. Sin embargo, no existe nada entre ambos tipos a partir de lo cual se puede hacer una selección. De ahí que, seleccionar un valor crítico que esté basado en los posibles resultados generalmente significaría tener que fijar como nivel de significación uno diferente del que de ordinario se eligió. Además, para cada prueba se deberá seguir este procedimiento.

x	$P(x$ menor)
0	0.0005
1	0.0059
2	0.0327
3	0.1133
4	0.2744
5	0.5000
⋮	⋮

Como otra posibilidad, se puede observar simplemente que, si se rechaza en este caso, —eligiendo cuatro veces la cerveza embotellada— la probabilidad de rechazar una H_0 verdadera (de la tabla) es 0.2744, que excede a $\alpha = 0.05$. Por tanto, se debe aceptar H_0. No obstante, si únicamente un cliente elige la cerveza embotellada, la probabilidad acumulada sería lo suficientemente pequeña [$P(x \leq 1) = 0.0059$] como para que se rechazará H_0. Por tanto, se pueden formular las siguientes reglas generales:

> Al rechazar H_0 se obtiene una probabilidad de cola, menor que α o igual a dicho valor.
> Al aceptar H_0 se presenta una probabilidad de cola, mayor que α.

Cabe observar que si se utiliza una prueba de dos colas, se deberá sustituir con $\alpha/2$ el valor de α de las reglas anteriores. Además, se debe incluir, como parte del área de la cola, la probabilidad del resultado de la muestra. En otras palabras, si se encuentran 4 productos defectuosos en una muestra, el área de la cola es $P(x \leq 4)$ y no $P(x < 4)$.

EJERCICIOS

1. Para cada una de las siguientes condiciones represente gráficamente la distribución de muestreo apropiada (muestra grande), e indique en el esquema si la prueba es de una o de dos colas, y muestre la región en la(s) cola(s):
   a. $H_0: p = 35\%, H_1: p > 35\%, \alpha = 0.05$
   b. $H_0: p = 1\%, H_1: p < 1\%, \alpha = 0.05$
   c. $H_0: p = 22.5\%, H_1: p < 22.5\%, \alpha = 0.01$
   d. $H_0: p = 30\%, H_1: p \neq 30\%, \alpha = 0.05$
2. Un proveedor asegura que solamente el 2% de las piezas que surte fallarán en condiciones normales de trabajo. De una remesa de 5000 partes, se tomó al azar una muestra de 200 y se encontró que 10 tenían fallas.
   a. Establezca H_0 y H_1.
   b. ¿Estaría dispuesto a aceptar la afirmación del proveedor a un nivel 0.05?
3. Una moneda supuestamente justa se tira al aire 144 veces, 90 de las cuales cae cara. ¿Cree que la moneda es justa? Explíquelo.
4. Un senador asegura que un 20% o menos de los votantes de su Estado apoyan una ley que está siendo estudiada por un subcomité. Una encuesta entre 100 votantes tomados al azar indicó que 11 apoyarían la ley.
   a. Suponiendo que la aseveración del senador es verdadera, ¿cuántos votantes en favor deberían esperarse de una muestra de 100?
   b. Establezca H_1.
   c. Ponga a prueba esto al nivel 0.01.
5. El gobierno sostiene que 15% o menos de las familias de una determinada región, obtienen menos del salario mínimo establecido. Una muestra al azar de 60 familias dio como resultado 12 familias en tales condiciones. Pruebe la afirmación considerando $p > 15\%$.
6. Un artículo periodístico señala que 40% de los votantes, con edades que van de 18 a 21 años, no votaron en las últimas elecciones generales. Suponga que una muestra aleatoria de ocho votantes de este grupo es entrevistada obteniendo como resultado que dos de ellos no votaron. Pruebe $H_0 = 0.4$ con cada una de las siguientes alternativas al nivel 0.01:

   a. $H_1: p \neq 4$ b. $H_1: p > 0.4$ c. $H_1: p < 0.4$

7. Si un nuevo fármaco no tiene ningún efecto en los pacientes que la utilizan, algunos dirán que es útil, otros que se sienten peor, y otros más expresarán que se sienten igual. Si es efectivo, más personas asegurarían sentirse mejor; por el contrario, si tiene un efecto negativo, más individuos dirían sentirse peor. Suponga que de 20 que se administraron el fármaco, 11 dijeron sentirse mejor, 3 peor y 6 no haber experimentado ningún cambio. No tome en cuenta a estos últimos (es decir, utilice $n = 14$). Pruebe $H_0: p = 0.5$ respecto a las siguientes alternativas utilizando el nivel 0.05:
   a. El fármaco tiene efecto positivo.
   b. El medicamento tiene efecto negativo.

PRUEBA DE PROPORCIONES DE DOS MUESTRAS

El objeto de una prueba de dos muestras es determinar si las dos *muestras independientes* fueron tomadas de dos poblaciones, las cuales presentan la misma proporción de elementos con determinada característica. La prueba se concentra en la diferencia relativa (diferencia dividida entre la desviación estándar de la distribución de muestreo) entre las dos proporciones muestrales. Diferencias pequeñas denotan únicamente la variación casual producto del muestreo (se acepta H_0), en tanto que grandes diferencias significan lo contrario (se rechaza H_0). El valor estadístico de prueba (diferencia relativa) es comparado con un valor tabular de la distribución normal, a fin de decidir si H_0 es aceptada o rechazada. Una vez más, esta prueba se asemeja considerablemente a la prueba de medias de dos muestras.

La hipótesis nula en una prueba de dos muestras es

$$H_0: p_1 = p_2$$

Las hipótesis alternativas posibles son

$$H_1: p_1 \neq p_2 \qquad H_1: p_1 > p_2 \qquad H_1: p_1 < p_2$$

Como de costumbre, el método que se debe seguir consiste en suponer inicialmente que H_0 es verdadera y después utilizar una distribución de muestreo, basada en ese supuesto, a fin de llevar a cabo la prueba. Sin embargo, a diferencia de la prueba de una muestra, no existe ningún parámetro de población establecido en H_0. Por tanto, la obtención de un valor de p, a fin de emplearlo, debe ser manejado de una manera un tanto diferente. Obsérvese que si, de hecho, p_1 es igual a p_2, entonces las dos muestras obtenidas a partir de dos poblaciones, se pueden considerar como dos muestras tomadas de la *misma* población. De ese modo, cada proporción muestral puede considerarse como una estimación de la misma proporción de la población. Además, parecería razonable (debido al mayor tamaño) que la combinación de las dos muestras proporcionara una estimación aún mejor del verdadero valor de la proporción de población. La estimación combinada de p se puede calcular de la siguiente manera:

$$p = \frac{x_1 + x_2}{n_1 + n_2}$$

en la cual

x_1 = número de aciertos en la muestra 1
x_2 = número de aciertos en la muestra 2
n_1 = número de observaciones de la muestra 1
n_2 = número de observaciones de la muestra 2

Este valor de p se utiliza para calcular la desviación estándar de la proporción, que es semejante a las fórmulas anteriores, excepto en que ahora se debe "ponderar" por medio de los dos tamaños de muestra:

$$\sigma_p = \sqrt{p(1-p)[(1/n_1) + (1/n_2)]}$$

Ejemplo 5 Considérese este caso. A los votantes de dos ciudades se les pregunta si están a favor o en contra de una ley que está actualmente en estudio en la legislatura del estado, para proporcionar ropa a animales de granja. Para determinar si los votantes de las dos ciudades difieren en términos del porcentaje que está a favor, se toma una muestra de 100 votantes de cada ciudad. Treinta de los muestreados de una ciudad están a favor, en tanto que, en la otra, lo están veinte.

En primer lugar se deben establecer las hipótesis nula y alternativa:

$$H_0: p_1 = p_2$$
$$H_1: p_1 \neq p_2$$
$$\alpha = .01$$

(Se deberá utilizar una prueba de dos colas, ya que el problema no especifica si se considera que el porcentaje de una ciudad debe ser mayor —o más alto— que el porcentaje en la otra.)

El valor estadístico de prueba z es

$$z = \frac{(x_1/n_1) - (x_2/n_2)}{\sqrt{p(1-p)[(1/n_1) + (1/n_2)]}} \quad \text{donde} \quad p = \frac{x_1 + x_2}{n_1 + n_2}$$

Por consiguiente, se tiene

$$p = \frac{30 + 20}{100 + 100} = 0.25$$

$$z = \frac{0.30 - 0.20}{\sqrt{0.25(0.75)[1/100) + (1/100)]}} = \frac{0.10}{\sqrt{0.00375}} = \frac{0.10}{0.06} = 1.63$$

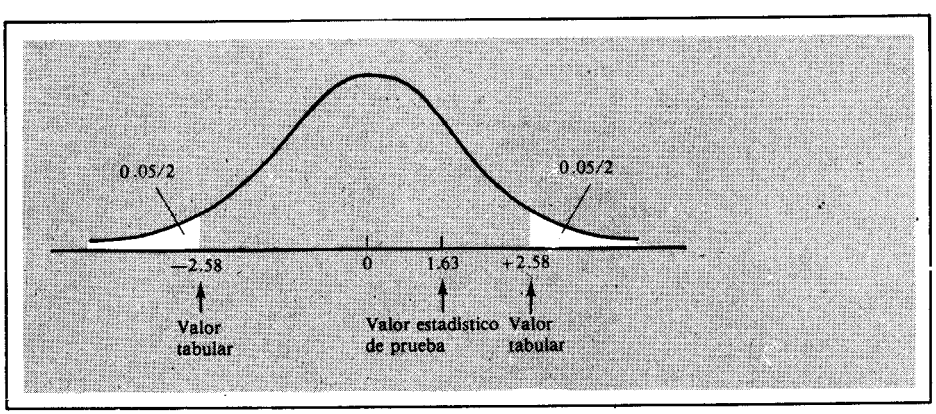

FIGURA 12.5 H_0 es aceptada.

Como el valor estadístico de prueba, z, presenta un valor dentro de la región de aceptación, no se puede concluir que las dos ciudades difieren en términos del porcentaje de quienes están a favor de la aprobación de la ley. Ver figura 12.5.

Ejemplo 6 Supóngase que, en el ejemplo anterior, se llegó a la hipótesis (H_1) de que p_1 era mayor que p_2. La prueba se podría establecer de la siguiente manera:

$$H_0: p_1 = p_2$$
$$H_1: p_1 > p_2$$
$$\alpha = 0.05$$

Utilizando los mismos valores muestrales, z permanecería igual a 1.63. De ahí que se aceptaría H_0 al nivel 0.05, y *no se podría* concluir que la primera ciudad presenta un porcentaje mayor de votantes a favor de la promulgación de la ley. Ver figura 12.6.

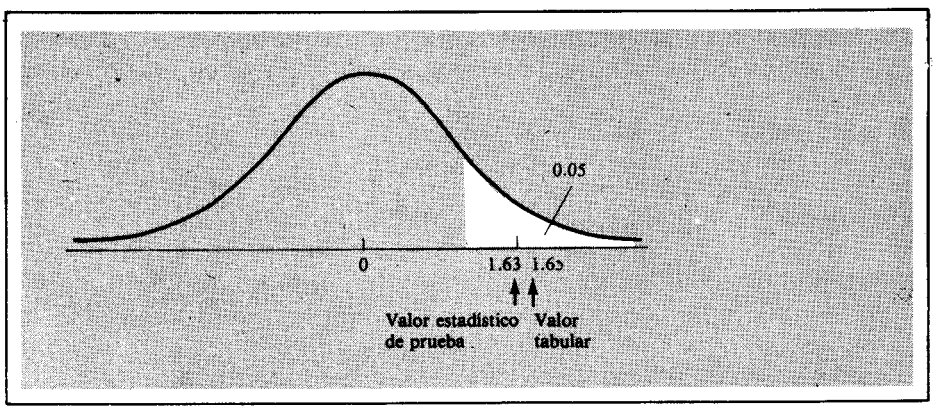

FIGURA 12.6 H_0 es aceptada.

Ejemplo 7 Supóngase que en el ejemplo 6, la alternativa fue que $P_1 < P_2$. Gráficamente, la prueba se indicaría como en la figura 12.7. Como el valor estadístico de prueba z es positivo, una vez más se aceptaría H_0.

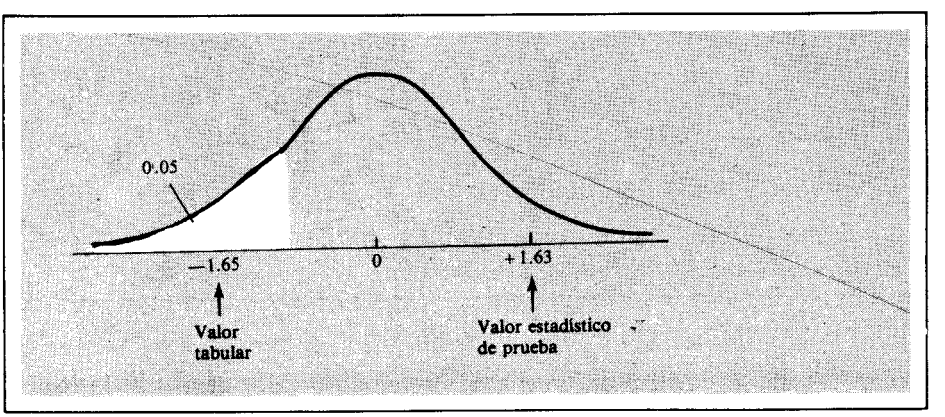

FIGURA 12.7 H_0 es aceptada.

EJERCICIOS

1. Un fabricante quiere saber si un recubrimiento especial que se aplica a la superficie de las placas de automóvil las hará más resistentes a la oxidación. Se distribuyen 20 000 placas recubiertas, junto con 20 000 placas no recubiertas. Una muestra aleatoria de placas recubiertas tomada un año después indica que 360 de 400 están en excelentes condiciones. Una muestra aleatoria de 225 placas no recubiertas indica que 195 de ellas se encuentran en excelentes condiciones. ¿Es posible concluir a partir de esto que cierto tipo de placas es superior al otro?
2. Una muestra al azar de 200 votantes registrados en dos distritos indicó que 45 de 200 votantes urbanos están a favor del reavalúo de la propiedad mientras que 55 votantes suburbanos de un grupo de 200, están en favor de la misma propuesta. ¿Se podría decir que la proporción de votantes que están en pro es la misma en estos distritos?
3. Se ponen a prueba dos métodos posibles para tapar botellas. En un lote de 1000, la máquina A genera 30 rechazos, en tanto que la máquina B solamente genera 20. ¿Es posible concluir (0.05) que las dos máquinas son diferentes?

PRUEBA DE PROPORCIONES DE k MUESTRAS

La finalidad de una prueba de k muestras es evaluar la aseveración que establece que todas las k muestras independientes provienen de poblaciones que presentan la misma proporción de algún elemento. De acuerdo con esto, las hipótesis nula y alternativa son

H_0: Todas las proporciones de la población son iguales.
H_1: No todas las proporciones de la población son iguales.

Considérese por ejemplo el siguiente caso. Un centro comercial compró y plantó 720 bulbos de tulipán de cuatro colores, para sus jardines:

blancos	200
rojos	160
amarillos	240
morados	120
	720

Por desgracia, no todos florecieron. El centro comercial quiere determinar si los "fracasos" eran *independientes* del color (es decir, si todas las proporciones de la población son iguales) antes de comprar más bulbos de tulipán.

Cada color se puede considerar como una población, y los bulbos de cada color, como una muestra de cada una de las poblaciones. Los resultados muestrales (florecer respecto de no florecer) se pueden obtener a partir de datos proporcionados por el jardinero. Los resultados han sido ordenados en una tabla de 2 por k ($2 \times k$): 2 filas y k columnas, una columna para cada muestra. Las k muestras se listan en hileras, y los resultados muestrales en columnas:

	Muestra				
Resultados muestrales	Blanco	Rojo	Amarillo	Morado	Totales
florecieron	176	136	222	114	648
no florecieron	24	24	18	6	72
total de plantados	200	160	240	120	720

Si, de hecho, la hipótesis nula es verdadera, entonces las variaciones entre las muestras se deben únicamente al azar. Supóngase inicialmente que este es el caso (es decir, que H_0 es verdadera). Entonces las cuatro muestras se deben considerar como cuatro muestras provenientes de la misma población. Al combinar los resultados de las muestras, se puede obtener una estimación de la proporción de la población verdadera de bulbos que tienden a florecer. De acuerdo con lo anterior, se observa que

$$p = \frac{176 + 136 + 222 + 114}{200 + 160 + 240 + 120} = 90\%$$

Esta estimación del porcentaje de la población se puede utilizar para calcular el número esperado de "éxitos" de cada categoría bajo el supuesto de que la hipótesis nula es verdadera. Estos números esperados servirán a su vez como base respecto a la cual se puedan evaluar los resultados observados (muestra). Si las diferencias existentes entre los dos resultados son pequeñas, la conclusión de que éstas se pueden deber a la casualidad parecería razonable. Sin embargo, las grandes diferencias parecerían indicar que el índice de sobrevivencia de los bulbos difiere en cada uno de los distintos colores de bulbos.

El número esperado de éxitos para cada categoría se puede determinar multiplicando el número total de bulbos plantados por el porcentaje estimado de la población:

$$p \times \text{cantidad plantada} = \text{cantidad esperada}$$

Así,

blanco:	$90\% \times 200 = 180$
rojo:	$90\% \times 160 = 144$
amarillo:	$90\% \times 240 = 216$
morado:	$90\% \times 120 = 108$

El número esperado de fracasos de cada color también se podría determinar de la misma manera (es decir, obteniendo primero el porcentaje del total de los que fallan y luego multiplicando dicho porcentaje por el número de flores de cada color). Sin embargo, como se conoce el número esperado de éxitos de cada color, al igual que el número total de flores plantadas, una simple substracción indicará el número esperado de fracasos:

Color	Número de flores plantadas	− Florecimientos esperados	= Fracasos esperados
blanco:	200	180	20
rojo:	160	144	16
amarillo	240	216	24
morado	120	108	12

A menudo es conveniente incluir tanto las frecuencias esperadas como las frecuencias observadas en una sola tabla, para propósitos de análisis. Las frecuencias esperadas generalmente se indican entre paréntesis.

	Blanco	Rojo	Amarillo	Morado
Florearon	(180)	(144)	(216)	(108)
	176	136	222	114
Fallaron	(20)	(16)	(24)	(12)
	24	24	18	6

Es importante que la frecuencia *esperada* para cada casilla sea por lo menos de 5, a fin de que sea válida la técnica presentada aquí. Si no se cumple esta condición, se deberán agrupar una o más hileras (o columnas), para obtener la frecuencia esperada mínima. En este ejemplo, la menor frecuencia esperada es 12, de manera que no es necesario preocuparnos por este problema. Sin embargo, si por ejemplo, las últimas dos columnas tienen frecuencias esperadas bajas en ciertas casillas, se podría considerar la categoría "amarillo y morado" como una sola columna, y combinar las frecuencias esperadas y observadas.

La diferencia entre los dos conjuntos de frecuencias se puede determinar calculando el siguiente valor estadístico de prueba:

$$\chi^2 = \sum \left[\frac{(\text{frecuencia observada} - \text{frecuencia esperada})^2}{\text{frecuencia esperada}} \right]$$

Por procedimiento, esto equivale a 1) restar la frecuencia esperada de la frecuencia observada para cada casilla; 2) elevar al cuadrado cada una de estas diferencias; 3) "estandarizar" las diferencias elevadas al cuadrado dividiendo cada una entre la frecuencia esperada de cada casilla; y finalmente, 4) sumar todos los términos para obtener el valor total. El total recibe el nombre de valor *estadístico* de *prueba ji* o *chi cuadrada* (χ^2). Por tanto, el valor estadístico de prueba χ^2 para este problema se calcularía de la siguiente manera:

$$\chi^2 = \frac{(176-180)^2}{180} + \frac{(136-144)^2}{144} + \frac{(222-216)^2}{216} + \frac{(114-108)^2}{108}$$
$$+ \frac{(24-20)^2}{20} + \frac{(24-16)^2}{16} + \frac{(18-24)^2}{24} + \frac{(6-12)^2}{12}$$
$$= 0.09 + 0.44 + 0.17 + 0.33 + 0.80 + 4.00 + 1.50 + 3.00$$
$$= 10.33$$

Este valor estadístico de prueba se debe comparar con un valor crítico (tabular) de la distribución de muestreo ji cuadrada, a fin de decidir si se acepta o rechaza H_0.

Distribución de muestreo ji cuadrada

Como sucede con las distribuciones t y F, la distribución ji cuadrada tiene una forma que depende del número de grados de libertad asociados a un determinado problema. Varias de estas curvas se ilustran en la figura 12.8. Debido a esta tendencia, el valor crítico (valor que deja un determinado porcentaje de área en la cola) será función de los grados de libertad. Así, para obtener un valor crítico a partir de una tabla de ji cuadrada, se debe seleccionar un nivel de significación y determinar los grados de libertad para el problema que se esté analizando.

Los grados de libertad son una función del número de casillas en una tabla de $2 \times k$. Es decir, los grados de libertad reflejan el tamaño de la tabla. Cabe recordar que, al calcular las frecuencias esperadas, tanto para las columnas como para las filas,

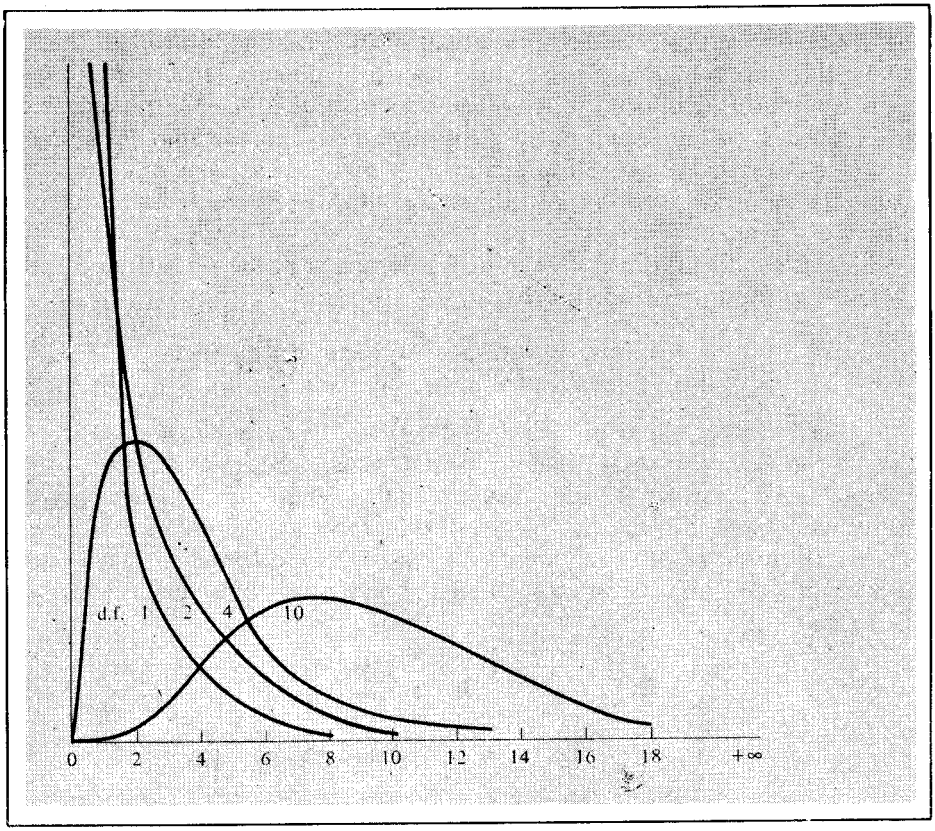

FIGURA 12.8 La forma de la distribución ji cuadrada depende de sus grados de libertad.

el valor esperado de la última celda se podría obtener simplemente restando la suma de las otras frecuencias esperadas de esa fila o columna, de las filas o columnas totales. De este modo, las frecuencias esperadas se ven restringidas por el requisito de que se suman al total, y se utiliza este conocimiento para obtener las frecuencias finales. Para compensar esto, se dice que cada columna pierde un grado de libertad. Así, los grados de libertad de la columna son el número de filas (categorías) menos 1, o bien, $r - 1$. En forma similar, como también se conocían los totales de las hileras, podría obtenerse cualquier valor esperado de cada una de ellas utilizando la diferencia entre el total de la fila y la suma de las otras frecuencias de la misma. Por consiguiente, cada fila presenta un número de grados de libertad igual al número de columnas (muestras) menos 1, o bien, $k - 1$. El efecto neto es que el número de grados de libertad para la tabla es el producto de (número de filas -1) $\times$ (número de columnas -1), o bien, $(r - 1)(k - 1)$. Por tanto, con 2 hileras y 4 columnas, los grados de libertad son igual a $(2 - 1)(4 - 1) = 3$.

En la tabla I del Apéndice se proporcionan valores críticos de ji cuadrada para diversos grados de libertad en el caso de niveles de significación seleccionados.

La prueba requiere la comparación del valor calculado de ji cuadrada con el obtenido a partir de una tabla de valores críticos de esta última, utilizando los grados de libertad apropiados. Si el valor estadístico de prueba es menor que el valor tabular, la hipótesis nula es aceptada; si no ocurre así, H_0 es rechazada. Esto se ilustra en la figura 12.9. (Cabe observar que un valor de ji cuadrada nunca puede ser menor que cero, dado que se calcula utilizando desviaciones elevadas al cuadrado.)

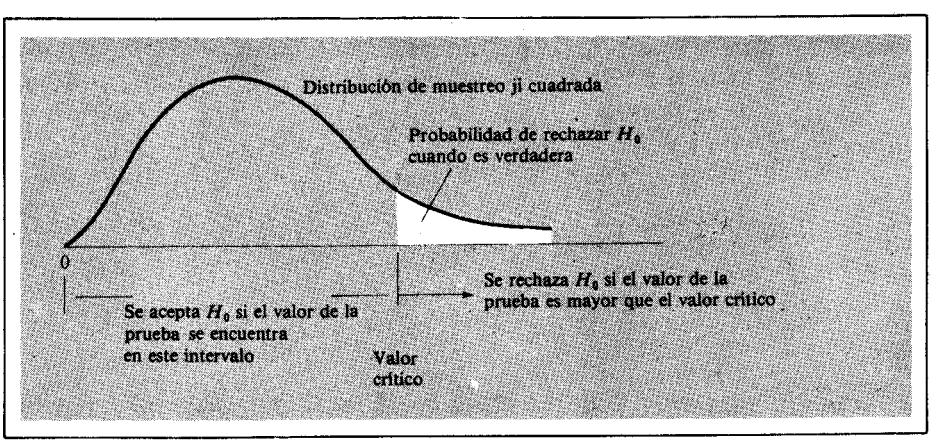

FIGURA 12.9 Un valor estadístico de prueba ji cuadrada menor que el valor crítico o igual a él se considera como prueba de la variación casual; H_0 es aceptada.

Si se compara el valor calculado de χ^2 con uno que se obtuvo a partir de la tabla de valores críticos (ver tabla I del Apéndice), utilizando, por ejemplo, el nivel 0.01, se observa que la variación entre las proporciones muestrales se puede deber al azar, ya que el valor estadístico de prueba (10.33) cae dentro de la región de aceptación. Por tanto,

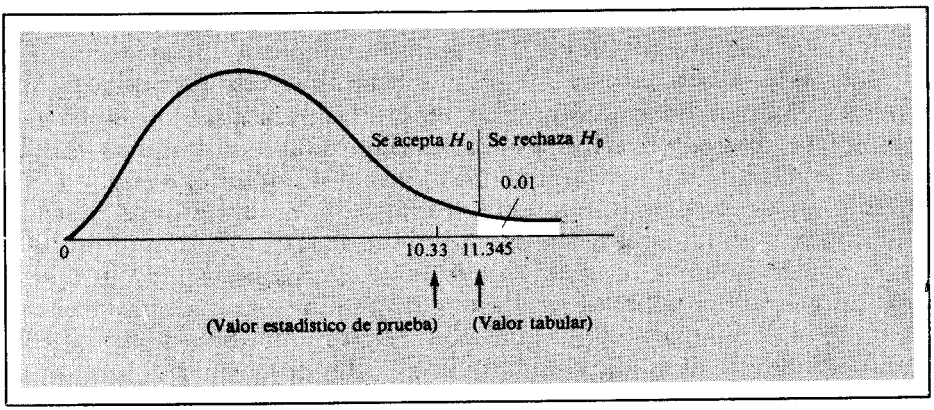

FIGURA 12.10 Se acepta la hipótesis nula al nivel 0.01.

no se puede concluir que las proporciones de la población no sean todas iguales. La prueba se ilustra en la figura 12.10.

Análisis de una tabla de r por k

El análisis de una tabla de r por k (esto es, $r \times k$) es una extensión del análisis de una tabla de $2 \times k$. Cada tabla todavía tiene k columnas (muestras) pero ahora hay más de dos hileras. Esto significa que los resultados de la muestra son clasificados *en más de dos* categorías. De este modo, las poblaciones son consideradas como *multinomiales*. El formato de la tabla se muestra a continuación:

		Muestras				
		1	2	3	···	k
categorías muestrales	1					
	2					
	3					
	⋮					
	r					

La ventaja de que haya más de dos clases por muestra es que esto proporciona una diferenciación más fina que facilita la comparación; cuanto más fina sea dicha diferenciación, mejores serán las posibilidades de distinguir entre muestras de poblaciones con proporciones iguales y las de proporciones diferentes. Las filas adicionales no presentan ningún cambio en el procedimiento de cálculo, excepto porque intervienen más casillas.

Como en la prueba $2 \times k$, las hipótesis nulas y alternativas son

H_0: Todas las proporciones de la población son iguales.
H_1: No todas las proporciones de la población son iguales.

Considérese este ejemplo. En un estudio reciente para determinar si la preferencia por determinados sabores de helados varía en diferentes partes del país, se recopilaron los siguientes datos. Determine si es posible comparar las tres regiones, utilizando el nivel de significación 0.05.

Frecuencias observadas

	Ubicación			
Sabor del helado	Northeast	Sur	Oeste	Totales
Vainilla	86	44	70	200
chocolate	45	30	50	125
Fresa	34	6	10	50
Otros	85	20	20	125
Totales	250	100	150	500

Si la hipótesis nula es aceptada, esto indica que la preferencia por cierto sabor es *independiente* de la región; si se rechaza H_0, esto quiere decir que la preferencia de sabor *depende* de la región. De este modo, las hipótesis nula y alternativa también se pueden enunciar en los siguientes términos:

H_0: La preferencia por cierto sabor es independiente del lugar (región).
H_1: La preferencia por cierto sabor depende de (está relacionada con) el lugar.

La hipótesis nula se puede interpretar de la siguiente manera: todos los porcentajes de cada población en la categoría 1 son iguales; los porcentajes de cada población en la categoría 2 son todos iguales; todos los porcentajes de cada población en la categoría 3 son iguales; y así sucesivamente. Es decir,

		Población							
		1		2		3	$\cdots$	k	
categoría	1	$p_{1,1}$	$=$	$p_{1,2}$	$=$	$p_{1,3}$	$=$	$p_{1,k}$	
	2	$p_{2,1}$	$=$	$p_{2,2}$	$=$	$p_{2,3}$	$=$	$p_{2,k}$	
	$\vdots$	$\vdots$		$\vdots$		$\vdots$		$\vdots$	
	r	$p_{r,1}$	$=$	$p_{r,2}$	$=$	$p_{r,3}$	$=$	$p_{r,k}$	

Una vez más, el procedimiento de la prueba comprende la determinación de las frecuencias esperadas de casilla, bajo el supuesto de que H_0 es verdadera y el cálculo de un valor estadístico de prueba que refleja las desviaciones elevadas al cuadrado entre cada par de frecuencias de casillas observadas y esperadas. Para obtener las frecuencias esperadas, se pueden utilizar los porcentajes de las hileras, al igual que los de las columnas, los resultados serán iguales en ambos casos; la selección de cuál utilizar a menudo depende de con cuáles frecuencias es más fácil trabajar (los totales de las filas o los totales de las columnas).

Los porcentajes de las hileras se determinan calculando la razón del total de las hileras respecto al número total de observaciones:

$$p_{\text{fila}} = \frac{\text{total de la hilera}}{\text{número total de observaciones}}$$

En forma similar, cada porcentaje de columna se puede calcular como

$$p_{\text{col}} = \frac{\text{total de la columna}}{\text{número total de observaciones}}$$

Por tanto, las frecuencias esperadas de cada casilla individual se determinan calculando

$$P_{\text{fila}} \times \text{total de la columna, } \textit{o bien, } P_{\text{col}} \times \text{total de la fila}$$

La suma de las desviaciones de la casilla dependerá en cierto modo del tamaño de la tabla, lo cual se refleja mediante los grados de libertad:

$$\text{grados de libertad} = (r - 1)(k - 1)$$

donde

r = número de filas
k = número de columnas

Los cálculos respecto de los datos acerca de los sabores de helado se utilizan como se indica a continuación: Primero se determinan los porcentajes de las filas:

$$\text{Fila 1 (vainilla):} \quad \frac{200}{500} = 0.40 = p_1$$

$$\text{Fila 2 (chocolate):} \quad \frac{125}{500} = 0.25 = p_2$$

$$\text{Fila 3 (fresa):} \quad \frac{50}{500} = 0.10 = p_3$$

$$\text{Fila 4 (otros):} \quad \frac{125}{500} = 0.25 = p_4$$

Posteriormente se utilizan los porcentajes de las filas para obtener las frecuencias esperadas para cada casilla.

Sabores	Número esperado de respuestas: Noreste	Sur	Oeste
vainilla	$250 \times 0.40 = 100$	$100 \times 0.40 = 40$	$150 \times 0.40 = 60$
chocolate	$250 \times 0.25 = 62.5$	$100 \times 0.25 = 25$	$150 \times 0.25 = 37.5$
fresa	$250 \times 0.10 = 25$	$100 \times 0.10 = 10$	$150 \times 0.10 = 15$
otros	$250 \times 0.25 = 62.5$	$100 \times 0.25 = 25$	$150 \times 0.25 = 37.5$
Total	250.0	100	150.0

Luego se calcula el valor estadístico de prueba,

$$\chi^2 = \sum\left[\frac{(o-e)^2}{e}\right]$$

$$\left.\begin{aligned}
&\frac{(86-100)^2}{100} + \frac{(44-40)^2}{40} + \frac{(70-60)^2}{60} = \frac{196}{100} + \frac{16}{40} + \frac{100}{60} = 1.96 + 0.40 + 1.67 \\
&\frac{(45-62.5)^2}{62.5} + \frac{(30-25)^2}{25} + \frac{(50-37.5)^2}{37.5} = \frac{306.25}{62.5} + \frac{25}{25} + \frac{156.25}{37.5} = 4.90 + 1.00 + 4.17 \\
&\frac{(34-25)^2}{25} + \frac{(6-10)^2}{10} + \frac{(10-15)^2}{15} = \frac{81}{25} + \frac{16}{10} + \frac{25}{15} = 3.24 + 1.60 + 1.67 \\
&\frac{(85-62.5)^2}{62.5} + \frac{(20-25)^2}{25} + \frac{(20-37.5)^2}{37.5} = \frac{506.25}{62.5} + \frac{25}{25} + \frac{306.25}{37.5} = 8.10 + 1.00 + 8.17
\end{aligned}\right\} = 37.88$$

A continuación se determina el valor tabular. Los grados de libertad son $(r - 1)(k - 1) = (4 - 1)(3 - 1) = 6$. El valor tabular es 12.592 (ver tabla I en el Apéndice). Finalmente, se compara el valor estadístico de prueba con el valor tabular (ver figura 12.11). Como el valor estadístico de prueba se encuentra en la región de rechazo, el estudio parece indicar que las regiones no son comparables en términos de

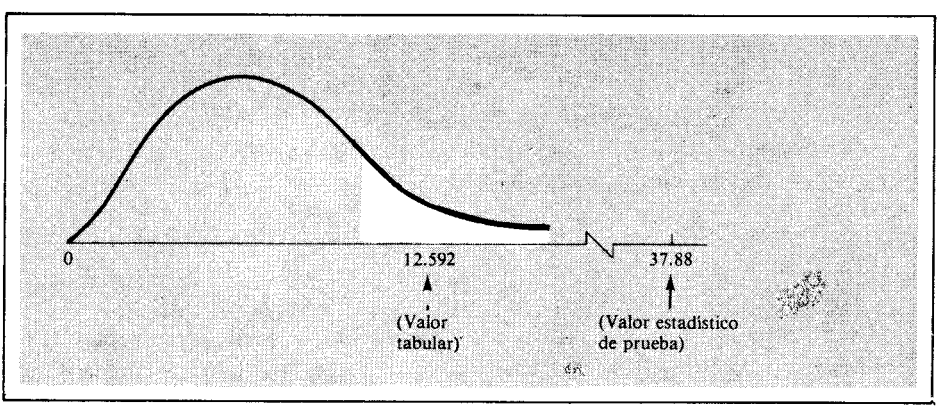

FIGURA 12.11 Comparación del valor estadístico de prueba con el valor tabular que indica un rechazo de H_0.

la preferencia por determinado sabor. En otras palabras, dicha preferencia parece *depender* de la región.

EJERCICIOS

1. Cada uno de los siguientes valores representa el tamaño de una tabla $r \times k$;

$$3 \times 4 \quad 4 \times 3 \quad 5 \times 5 \quad 2 \times 5 \quad 3 \times 6 \quad 4 \times 6$$

   a. Determine el número de grados de libertad para cada uno.
   b. Utilice una tabla ji cuadrada para obtener valores críticos en los niveles 0.01 y 0.05 para cada uno.

2. Determine las frecuencias esperadas para esta tabla de $r \times k$:

					Renglones totales
					200
					200
					200
Columnas totales	90	120	150	240	600

3. En un estudio para determinar si cuatro compañías constructoras son similares en términos del tipo de quejas expresadas por los nuevos propietarios, se toma una muestra aleatoria de 100 casas construidas por cada una de las compañías, registrándose la queja principal de los propietarios. ¿Sugieren los resultados (0.05) que los constructores son semejantes?

Major principal	Constructor A	B	C	D
estructura	12	10	33	5
calefacción, drenaje, instalación eléctrica	28	5	17	30
orientación	40	60	20	40
otros	20	25	30	25
	100	100	100	100

4. En un estudio de facturas o notas de reparación, se incluyeron cinco centros de servicio para autos. Los siguientes datos representan una parte de esa investigación, que se concentró en los cargos de mano de obra como un porcentaje de la facturación total. ¿Cabría decir que los cinco centros son similares en este aspecto? Utilice el nivel 0.05.

Mano de obra (como porcentaje) de la facturación total	Centro de servicio de autos *A*	*B*	*C*	*D*	*E*	
0 a 15	10	5	8	9	13	45
15.1 a 25	10	8	14	10	18	60
25.1 a 35	18	13	4	20	50	105
35.1 a 50	22	4	34	11	19	90
número de autos	60	30	60	50	100	300

5. A una muestra aleatoria de 50 poseedores de boletos para toda la temporada de fútbol y 50 poseedores de boletos para un solo juego se les preguntó acerca del acierto de la distribución de localidades para los partidos con los siguientes resultados:

	Poseedores de boletos de temporada	Poseedores de boletos para un juego
La aprobaron	35	25
No la aprobaron	15	25

   a. Ponga a prueba estos resultados al nivel 0.05, utilizando un análisis $r \times k$. ¿Qué se puede concluir?
   b. Ponga a prueba estos resultados utilizando una prueba z de dos muestras, empleando el mismo nivel. ¿Qué se puede concluir?
   c. Eleve al cuadrado los valores de prueba y los de la tabla z, y compárelos con los valores de prueba ji cuadrada y los de la tabla.

6. A cuatro muestras de 30 alumnos se les pide que emitan comentarios acerca de las disposiciones escolares para el estacionamiento de automóviles. Valiéndose del nivel 0.01, ¿qué se puede concluir?

	De 1er grado	De 2° grado	De 3er grado	De 4° grado
Las aprobaron	5	4	20	27
Las desaprobaron	25	26	10	3

PRUEBA χ^2 DE BONDAD DE AJUSTE*

Este tipo de pruebas es utilizado para evaluar las afirmaciones con respecto a la distribución de valores en una población. Estas pruebas pueden servir para satisfacer una gran variedad de necesidades. Por ejemplo, muchos procedimientos estadísticos son válidos únicamente con determinados tipos de poblaciones (por ejemplo, las pruebas de medias de muestras pequeñas requieren poblaciones normales). En consecuencia, es una ventaja tener una forma de determinar si una población tiene la distribución re-

*Esta sección se puede omitir.

querida. De este modo, si existe alguna duda razonable con respecto a la población, puede ser prudente llevar a cabo una comprobación de la distribución antes de continuar.

Un segundo uso de una prueba de bondad de ajuste es determinar si tres o más categorías de una población son igualmente probables (una distribución multinomial). Otro uso es en situaciones en las que el método para resolver un problema puede diferir dependiendo del tipo de distribución que se esté tratando. Por ejemplo, el método para problemas que se presentan en las líneas de espera o colas (como las que se forman en las cajas de un supermercado, parada de autobuses o gasolineras) depende de la distribución de la llegada de los clientes, así como de la distribución de los tiempos de espera. Si se puede mostrar que dichas distribuciones son de cierta forma, podrán aplicarse al problema ecuaciones de tipo normal. Si no es así, se puede requerir de métodos alternativos, que generalmente son más rigurosos (es decir, simulaciones u operaciones matemáticas complejas).

La prueba χ_2 de bondad de ajuste es una variante de la prueba χ^2 que se estudió en la sección anterior. El cálculo del valor estadístico de prueba y su evaluación son muy semejantes en ambos casos, aunque existen unas cuantas excepciones. Entre las principales excepciones se encuentran la forma como se plantea H_0 y H_1, cómo se calculan las frecuencias esperadas y cómo se determinan los grados de libertad.

Como las pruebas de bondad de ajuste comprenden distribuciones, las hipótesis nula y alternativa deben especificar necesariamente un tipo de distribución. Además, la prueba para una distribución se puede concentrar simplemente en un tipo determinado (es decir, normal), o bien, en un tipo más parámetros (es decir, normal con una media de 5.2 y una desviación estándar de 2.4). Así, una hipótesis nula típica podría ser

H_0: La distribución de la población es de Poisson.

Esto proporciona solamente el tipo de distribución y no indica nada con respecto a los parámetros de la distribución. O bien, mediante H_0 se puede ser más explícito, como a continuación:

H_0: La población es de Poisson, con una media de 3.2

La importancia de esta diferencia radica en que, cuando se establecen los parámetros de distribución, se cuenta con información suficiente para poder hacer directamente una tabla de probabilidad (en este ejemplo, una tabla de Poisson) y obtener frecuencias esperadas relativas. Sin embargo, una hipótesis del primer tipo requiere información adicional con respecto a la media de la población y, como este valor no ha sido establecido, se tendrán que utilizar los datos de la muestra para estimar la media de la población. Así la completitud o falta de completitud de H_0 influirá en el hecho de que los valores estadísticos de muestra deban ser calculados primero o si se debe pasar directamente a una tabla de probabilidad, para obtener las frecuencias esperadas. Además, el número de grados de libertad también se relacionará con el hecho de que H_0 esté completa: cada valor estadístico de muestra que se utiliza para obtener frecuencias esperadas produce la pérdida de un grado de libertad. De este modo, si se debe calcular el valor medio de una

muestra, se perderá un grado de libertad; si se debe calcular la media y la desviación estándar, se perderán dos.

En realidad, una prueba de bondad de ajuste es una prueba de una muestra, pero en la que la población se ha dividido en k proporciones. Así pues, difiere de una prueba de proporciones de una muestra, estudiada anteriormente, la cual incluye sólo dos categorías (éxito y fracaso) en la población.

Por ejemplo, póngase a prueba un dado para verificar si no está cargado. Es decir, se quiere probar que

H_0: El dado no está cargado.
H_1: El dado está cargado.

La hipótesis nula no establece explícitamente parámetros de población, no obstante, por la naturaleza del problema, se sabe que, si el dado no está cargado, cabe esperar que la frecuencia de ocurrencias de cada uno de los seis posibles resultados (categorías) sean igualmente probables. Es decir, se pueden determinar las frecuencias esperadas en forma directa, sin tener que utilizar los datos de la muestra:

Categoría	Frecuencia esperada
	$\frac{1}{6} \times n$
	$\frac{1}{6} \times n$
	$\frac{1}{6} \times n$
	$\frac{1}{6} \times n$
	$\frac{1}{6} \times n$
	$\frac{1}{6} \times n$

En consecuencia, si se tira el dado, digamos, 180 veces, se puede *esperar* que $\frac{1}{6}$ de los resultados correspondan a cada una de las categorías. Por tanto, para cada categoría, la frecuencia esperada sería $\frac{1}{6}$ (180) = 30.

Ahora, supóngase que el dado es tirado 180 veces, con los siguientes resultados:

Categoría	Frecuencia observada
1	20
2	35
3	25
4	35
5	32
6	33

Cabe observar que los valores varían hasta cierto punto del valor esperado de 30 en cada categoría. Es evidente que no se puede esperar que un dado no cargado produzca exactamente 30 resultados de cada cara, debido a la variación aleatoria. Sin embargo, es posible utilizar una prueba de bondad de ajuste para determinar si las diferencias se deben únicamente a la variación casual en el muestreo o si sugieren que el dado está cargado (es decir, que las categorías no son igualmente probables).

El primer paso que se debe seguir para resolver el problema es calcular el valor estadístico de prueba de ji cuadrada, que es

$$\chi^2 = \sum\left[\frac{(o-e)^2}{e}\right]$$

donde

o = frecuencia observada para cada categoría
e = frecuencia esperada para cada categoría

Utilizando la información anterior, se puede calcular el valor estadístico de prueba.

Cara	Frecuencia observada	Esperada ($\frac{1}{6} \times 180$)	Diferencia $(o-e)$	$(o-e)^2$	$(o-e)^2/e$
1	20	30	−10	100	3.33
2	35	30	5	25	0.83
3	25	30	−5	25	0.83
4	35	30	5	25	0.83
5	32	30	2	4	0.13
6	33	30	3	9	0.30
	180	180	0		6.25

Así $\chi_2 = 6.25$.

Grados de libertad

Para obtener un valor de ji cuadrada en la tabla, para fines de comparación, en primer lugar es necesario determinar los grados de libertad. Esto, a su vez, depende del número de *restricciones* comprendidas en el cálculo del valor estadístico de prueba. Como una regla general, los grados de libertad para una prueba de bondad de ajuste es igual al número de categorías k, menos el número de limitaciones presentes en los datos de la muestra. Siempre hay por lo menos una restricción: Las frecuencias esperadas se deben sumar al total. Se presentan mayores limitaciones si los valores estadísticos de muestra (media, desviación estándar, etc.) se utilizan para determinar las frecuencias esperadas. Se pierde un grado de libertad por cada valor estadístico de muestra empleado de esa manera. Por tanto, los grados de libertad para una prueba de bondad de ajuste son:

$$(k-1)-c$$

donde

k = número de categorías o clases

c = número de valores estadísticos de la muestra que se utilizan para determinar frecuencias esperadas

En el ejemplo del dado había una sola restricción: el total de las frecuencias tenía que sumar 180. Los grados de libertad, por tanto, son $k - 1$ (donde k es = número de categorías), o bien, $6 - 1 = 5$.

En ocasiones, las categorías se deben ordenar para obtener un nivel mínimo para las frecuencias esperadas. En tanto que aun existe cierta discrepancia entre los estadísticos, la regla que se habrá de seguir consiste en que ninguna frecuencia *esperada* deberá ser menor que 1.0, y que solamente una o dos deberán tener valores tan bajos. Una regla más conservadora es requerir frecuencias esperadas de por lo menos 5. Sin embargo, aquí se utilizará el valor mínimo 1.0. Esto implica que cuando las frecuencias esperadas son menores que el mínimo requerido, se deben combinar.

Evaluación del valor estadístico de prueba

Con 5 grados de libertad, el valor de la tabla al nivel 0.05 es 11.07. Como el valor estadístico de prueba en el caso del problema del dado (6.25) es menor que éste, se concluye que las diferencias entre las frecuencias observadas y esperadas se pueden deber, razonablemente, al azar. De este modo se acepta la afirmación que establece que un dado no está cargado,. Ver figura 12.12.

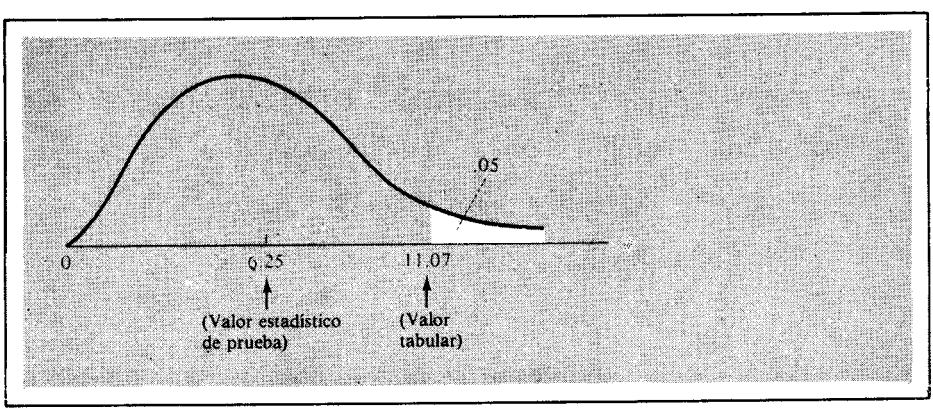

FIGURA 12.12 Se acepta H_0 al nivel de 0.05.

Ejemplo 8 Se dice que una máquina que llena y tapa envases de cerveza produce botellas con una media de un litro y una desviación estándar de 0.2 litros. Además, se afirma que la distribución de la cantidad de cerveza por botella está distribuida normalmente. Se examinan 100 botellas, y se registra la cantidad de cerveza por botella. Ponga a prueba esta aseveración utilizando el nivel de significación 0.25.

TABLA 12.1 Categoría y Datos para el ejemplo 8

Clase	Curva normal	Observada	Esperada (tabla normal × n)	$(o-e)$	$(o-e)^2$	$(o-e)^2/e$
Menor que 0.96	Menor que $\mu - 2\sigma$	4	$0.0228 \times 100 =$ 2.28	+1.72	2.958	1.30
0.96 a < 0.97	$\mu - 2\sigma$ a $< \mu - 1.5\sigma$	6	$0.0442 \times 100 =$ 4.42	+1.58	2.496	0.56
0.97 a < 0.98	$\mu - 1.5\sigma$ a $< \mu - 1\sigma$	4	$0.0919 \times 100 =$ 9.19	−5.19	26.936	2.93
0.98 a < 0.99	$\mu - 1\sigma$ a $< \mu - 0.5\sigma$	16	$0.1498 \times 100 =$ 14.98	+1.02	1.040	0.69
0.99 a < 1.00	$\mu - 0.5\sigma$ a < 0	20	$0.1915 \times 100 =$ 19.15	+0.85	0.722	0.04
1.00 a < 1.01	0 a $< \mu + 0.5\sigma$	18	$0.1915 \times 100 =$ 19.15	−1.15	1.322	0.07
1.01 a < 1.02	$\mu + 0.5\sigma$ a $< \mu + 1\sigma$	16	$0.1498 \times 100 =$ 14.98	+1.02	1.040	0.69
1.02 a < 1.03	$\mu + 1\sigma$ a $< \mu + 1.5\sigma$	10	$0.0919 \times 100 =$ 9.19	+0.81	0.656	0.07
1.03 a < 1.04	$\mu + 1.5\sigma$ a $< \mu + 2\sigma$	4	$0.0442 \times 100 =$ 4.42	−0.42	0.176	0.04
≥ 1.04	$\geq \mu + 2\sigma$	2	$0.0228 \times 100 =$ 2.28	−0.28	0.078	0.03
		100	100.00	0.00		$\chi^2 = 6.42$

Solución:

La hipótesis que se está probando es:

H_0: La distribución es normal, con una media de un litro y una desviación estándar de 0.2 litros.

La hipótesis alternativa es

H_1: La distribución no es normal, con una media de un litro y una desviación estándar de 0.2 litros.

La diferencia principal entre una prueba de bondad de ajuste que comprende una distribución continua, y otra que comprende una distribución discreta, radica en que no existen categorías naturales en el caso continuo. En consecuencia, se debe decidir cuantas categorías utilizar. Las principales consideraciones son las siguientes: 1) cuanto mayor sea el número de categorías, mayor será la oportunidad de descubrir las diferencias entre las frecuencias observadas y esperadas; y 2) las frecuencias esperadas no deberán caer por abajo de 1.0, y la mayoría de las frecuencias de las casillas deberán ser mayores que 5. Como un primer cálculo se deberá intentar utilizar entre 5 y 15 categorías. Como en este caso hay 100 observaciones, se puede intentar con 10 categorías.

Una posibilidad para seleccionar categorías (existen muchas) sería considerar categorías como las que se indican en la tabla 12.1. Las frecuencias esperadas se derivan de una tabla normal, y las frecuencias observadas se obtienen de los datos muestrales. Es decir, los valores observados son mediciones convertidas en una distribución de frecuencia que se conforma a las categorías presentadas. Cabe observar que los tamaños de las clases son diferentes. Las clases extremas han sido agrupadas.

El análisis de los datos revela que las diferencias observadas están dentro de lo que razonablemente se puede atribuir a la variación casual en el muestreo. Por tanto, se acepta la afirmación de una distribución normal con una media de un litro y una desviación estándar de 0.2 litros (ver figura 12.13).

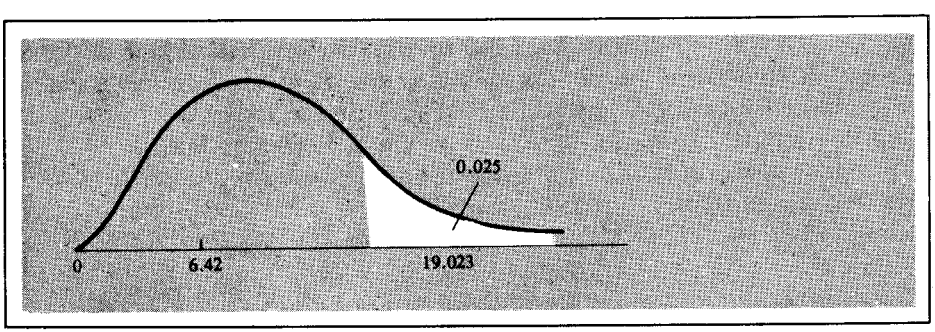

FIGURA 12.13 H_0 es aceptada.

Si se hubiera rechazado H_0, esto podría deberse a cualquiera de estos factores: 1) la media no era 1; 2) la desviación estándar no era 0.2; 3) la distribución no era normal; 4) alguna combinación de estos tres factores; o bien, 5) se cometió un error de tipo I. Se debe llevar a cabo un análisis ulterior para determinar qué razón es la más probable.

Empleo de datos muestrales para determinar frecuencias esperadas

Si las frecuencias esperadas no pueden ser determinadas sin la ayuda de los datos muestrales, se perderán grados de libertad adicionales a razón de un grado de libertad por cada valor estadístico de muestra así utilizado. Considérese este ejemplo, en el que se utiliza un valor estadístico de muestra: la media. Un ingeniero de tránsito quiere poner a prueba la afirmación de que los accidentes en cierto tramo —de dos kilómetros— de una autopista está distribuida mediante el método de Poisson. Un estudio acerca de los datos sobre accidentes proporciona las siguientes cifras.

Número de accidentes	Número de semanas
0	86
1	114
2	70
3	60
4	32
5	16
6	9
7	4
8	5
9	4
10 o más	0
	400

Las hipótesis nula y alternativa son

H_0: Los accidentes están distribuidos mediante el método de Poisson.

H_1: Los accidentes no están distribuidos mediante el método de Poisson.

Como la media de una distribución hipotética no está establecida (una distribución de Poisson está completamente especificada por su media), no se pueden obtener directamente las frecuencias relativas. La media de la población se debe estimar a partir de los datos de la muestra. Al hacerlo, se perderá un grado de libertad más.

La media de la distribución de frecuencias se puede encontrar utilizando la fórmula para los datos agrupados:

$$\bar{x} = \frac{\sum xf}{n}$$

en la cual

$$n = \sum f.$$

f frecuencia	x número	$x \cdot f$
0	86	0
1	114	114
2	70	140
3	60	180
4	32	128
5	16	80
6	9	54
7	4	28
8	5	40
9	4	36
	400	800

$$\text{media} = \frac{800}{400} = 2.0$$

Esta media de la muestra puede utilizarse junto con una tabla de probabilidades de Poisson, para obtener las frecuencias esperadas. Las probabilidades para cada resultado, 0, 1, 2, 3, etc., se leen directamente de la tabla, y cada uno es multiplicado por su frecuencia total. Por ejemplo, en la tabla de Poisson, $P(0) = 0.1353$. Multiplicando esto por 400, se obtiene una frecuencia esperada para 0 de 54.12. Los valores restantes se obtienen en forma semejante. En la tabla 12.2 se muestran los cálculos para este ejemplo. Las frecuencias esperadas se indican en la tercera columna. De la cuarta a la sexta columna se ilustran otros cálculos necesarios para obtener el valor estadístico de prueba de ji cuadrada.

Un requisito de la prueba de ji cuadrada es que las frecuencias *esperadas* para cada categoría deben ser 1 o más. Como las últimas dos frecuencias esperadas de la tabla 12.2 son menores de 1, se tendrán que combinar algunas de las categorías. Combinan-

TABLA 12.2 Datos y cálculos para el ejemplo de los accidentes en la autopista

Frecuencias observadas (número de accidentes)	Número de semanas	Frecuencias esperadas (Poisson $P \times 400$)	$(o - e)$	$(o - e)^2$	$(o - e)^2/e$
0	86	$0.1353 \times 400 = 54.12$	31.88	1016.36	18.78
1	114	$0.2707 \times 400 = 108.28$	5.72	32.72	0.30
2	70	$0.2707 \times 400 = 108.28$	−38.28	1465.36	13.53
3	60	$0.1804 \times 400 = 72.16$	−12.16	147.87	2.05
4	32	$0.0902 \times 400 = 36.08$	−4.08	16.65	0.46
5	16	$0.0361 \times 400 = 14.44$	1.56	2.43	0.17
6	9	$0.0120 \times 400 = 4.80$	4.20	17.64	3.68
7	4	$0.0034 \times 400 = 1.36$			
8	5 (13)	$0.0009 \times 400 = 0.36$ (1.80)	11.20	125.44	69.69
9	4	$0.0002 \times 400 = 0.08$			
	400	400.00*	0		108.66

*El total es aproximado debido al redondeo de las cantidades.

do las últimas dos, se obtiene un valor esperado de 0.44: 0.36 + 0.08 = 0.44. Incluso este valor es demasiado pequeño, lo que significa que se debe incluir por lo menos una categoría más. La siguiente frecuencia esperada, situada un peldaño arriba en la columna, es 1.36, que es suficiente para satisfacer la necesidad de una frecuencia esperada de 1.0 o más. Esto nos deja con 8 clases, y la última clase tiene una frecuencia esperada de 1.80: 1.36 + 0.36 + 0.08 = 1.80. Se debe desarrollar un agrupamiento semejante para las últimas tres frecuencias observadas, a fin de tener las correspondientes frecuencias esperadas y observadas. Así, la frecuencia observada de la última clase se convierte en 4 + 5 + 4 = 13.

A partir de la tabla 12.2 se observa que el valor estadístico de prueba es 108.66. Los grados de libertad se determinan mediante $(k - 1) - c$. Después de combinar las casillas, hay ocho categorías, por lo que $k = 8$. Además, se presentan dos limitaciones: el total tiene que equivaler a 400 y la media debe estimarse a partir de la muestra. Por tanto, $c = 1$. Así, los grados de libertad son $(8 - 1) - 1 = 6$.

Comparando el valor estadístico de prueba con los valores de la tabla a 6 grados de libertad, se observa que la hipótesis nula debería ser rechazada, ya que el valor estadístico de prueba es mayor que incluso el valor a 0.01, como se indica en la figura 12.14.

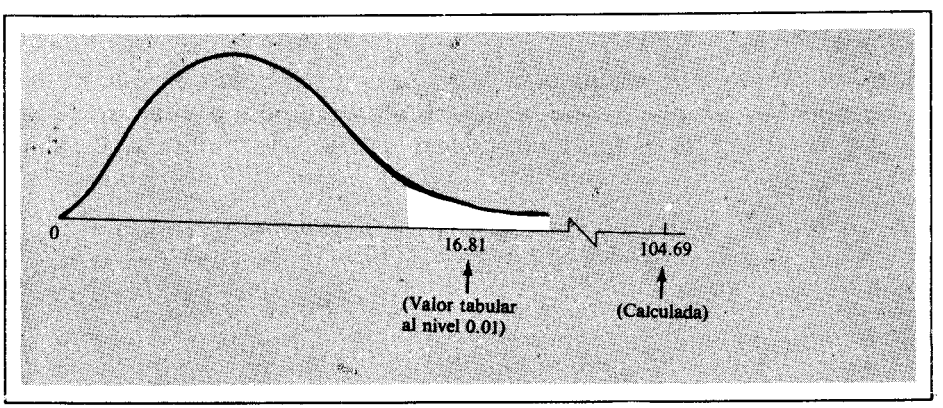

FIGURA 12.14 H_0 es rechazada.

EJERCICIOS

1. Utilice una prueba ji cuadrada de bondad de ajuste al nivel 0.05, para determinar cuál de las siguientes frecuencias muestrales están lo suficientemente próximas a las frecuencias esperadas, de manera que se puede aceptar la hipótesis nula.

a. Clase	Frecuencia observada	Frecuencia esperada
0	18	20
1	20	25
2	20	20
3	20	16
4	14	12
5	14	10
6	6	9
7	9	6
8	3	4
9	0	2
10	1	1
	125	125

b. Clase	Frecuencia observada	Frecuencia esperada
0	32	32.7
1	38	41.0
2	22	20.5
3	4	5.1
4	2	.6
5	2	.0
	100	100.0

c. Clase	Frecuencia observada	Frecuencia esperada
1–1.9	27	25
2–2.9	30	25
3–3.9	21	25
4–4.9	22	25
	100	100

2. En el curso de un año, en una compañía fabricante de artículos metálicos se han registrado 50 accidentes que han provocado retrasos en la producción. En una investigación realizada por el cuerpo de seguridad, el interés radicó en conocer el día de la semana en el que ocurre un accidente. A partir de los siguientes datos, ¿Cabría decir que el día de la semana constituye un factor importante? Ponga a prueba la hipótesis nula de que los días son igualmente probables.

Día	Número de accidentes
Lunes	15
Martes	6
Miércoles	4
Jueves	9
Viernes	16
	50

3. Un alumno diseñó un esquema para obtener dígitos aleatorios. Ha generado 1000 dígitos utilizando su método y quiere determinar si los dígitos son igualmente probables. Ponga a prueba sus datos utilizando un nivel de significación de 0.025.

											Total
Dígito	0	1	2	3	4	5	6	7	8	9	
Frecuencia	90	94	95	103	106	99	104	102	104	103	1000

4. La fabricación de tubería de acero requiere de una soldadura continua. Anteriormente se habían aproximado los defectos a lo largo de la soldadura de tubos de dos pulgadas, mediante una distribución de Poisson, con una media de 3 defectos/metro. Actualmente se está utilizando un nuevo tipo de máquina soldadora. Determine si el proceso ha cambiado.

												Total
Número de defectos	0	1	2	3	4	5	6	7	8	9	10 o más	
Frecuencia	5	14	16	20	18	17	3	2	4	0	1	100

5. Determine si los siguientes datos pueden provenir de un proceso de Poisson. Estime la media a partir de los datos muestrales.

								Total
Nacimiento/día	0	1	2	3	4	5	6 o más	
Número de días	3	11	4	2	2	1	2	25

6. Determine si las calificaciones de un considerable grupo de estudiantes de introducción a la psicología pueden ser aproximados mediante una distribución normal, con una media de 50.5 y una desviación estándar de 10. Utilice 0.05.

<25.5	14
26 a 30.5	18
31 a 35.5	22
36 a 40.5	20
41 a 45.5	40
46 a 50.5	30
51 a 55.5	22
56 a 60.5	20
61 a 65.5	2
66 a 70.5	6
71 a 75.5	—
76 a 80.5	4
≥ 81	2
	200

RESUMEN

Las pruebas de proporciones comprenden el análisis de los datos de conteo. La finalidad que persiguen tales pruebas es evaluar las afirmaciones respecto a proporciones de población. Una prueba de una muestra se utiliza para evaluar una proporción de una población especificada, mientras que una prueba de dos muestras estima la aseveración de que dos poblaciones tienen la misma proporción de algún elemento. Los procedimientos de prueba para una y dos muestras son muy semejantes para pruebas de medias de una y dos muestras.

Las pruebas de k muestras comprenden el uso de la distribución de muestreo ji cuadrada, que se asemeja en muchos aspectos a la distribución t. Las pruebas de k muestras se pueden utilizar para probar la igualdad de las proporciones de k poblaciones.

Las pruebas de bondad de ajuste son pruebas de una muestra que se utilizan para probar aseveraciones respecto a la forma de la distribución de una población.

TEMAS DE REPASO

1. ¿Cuál es el objeto general de una prueba de proporciones?
2. ¿Qué distribución de muestreo es adecuada para pruebas de una y dos muestras de proporciones cuando el tamaño de la muestra es pequeño? ¿Qué distribución de muestreo proporciona una aproximación que se puede utilizar con muestras grandes?
3. ¿Es necesario suponer que la población en una prueba de proporciones es aproximadamente normal? Explíquelo.
4. Establezca H_0 y H_1 en el caso de una tabla típica de $r \times k$.
5. ¿Qué ventaja hay en tener un conjunto de datos organizados en una tabla de $r \times k$, en lugar de una tabla de $2 \times k$? ¿Por qué, entonces, algunas veces es necesario utilizar una tabla de $2 \times k$?
6. ¿Cuál es la mínima frecuencia esperada, requerida para una tabla de ji cuadrada $r \times k$? ¿Para una prueba ji cuadrada de bondad de ajuste?
7. ¿En qué forma se parece la distribución ji cuadrada a la distribución t? ¿En qué forma difiere?
8. ¿Por qué una prueba ji cuadrada generalmente es una prueba de una cola?
9. ¿Cuáles son las dos pruebas ji cuadradas que se estudiaron en este capítulo?
10. ¿Cuál es la finalidad de una prueba de bondad de ajuste?

EJERCICIOS SUPLEMENTARIOS

1. Un conmutador telefónico está diseñado bajo la premisa de que las llamadas se registrarían a razón de 2.2 llamadas/minuto, y que las distribuciones de llamadas/minuto se podrían aproximar mediante la distribución de Poisson. Se llevan a cabo registros cuidadosos de la frecuencia de llamadas, en un lapso de 1000 minutos, con los siguientes resultados:

Llamadas/minuto	Número de minutos
0	20
1	130
2	150
3	220
4	185
5	140
6	100
7	55
8 o más	0
	1000

Utilice una prueba χ_2 al nivel 0.01, para probar la aseveración de Poisson con una media de 2.2 llamadas/minuto.

★2. De una tabulación cruzada de datos de un cuestionario se obtuvo la siguiente tabla:

Actitud hacia la política de defensa

	Republicanos	Demócratas	Independentes
La aprueban	35	80	50
No la aprueban	45	60	80
Se abstienen	20	60	70
	100	200	200

¿Muestra la tabla que la actitud hacia la política de defensa es independiente de la afiliación política? Utilice $\alpha = 0.01$.

3. Diga si el tiempo de entrenamiento se puede aproximar razonablemente mediante una distribución normal, con una media de 7 y una desviación estándar de 2 días para $\alpha = 0.05$, de acuerdo con los siguientes datos muestrales:

Tiempo	Frecuencia
2.01 - 3	1
3.01 - 4	3
4.01 - 5	4
5.01 - 6	10
6.01 - 7	15
7.01 - 8	9
8.01 - 9	4
9.01 - 10	1
10.01 - 11	2
11.01 - 12	1
	50

★4. Se cree que las llegadas de camiones a una gran estación están distribuidas mediante el método de Poisson, con una media de 2.4 llegadas/hora. Las observaciones se llevan a cabo en un periodo de 50 horas, con los siguientes resultados:

								Total
Número de camiones:	0	1	2	3	4	5	6	
Número de horas	5	10	15	10	5	2	3	50

a. Establezca H_0 y H_1.
b. ¿Qué se puede concluir con respecto a dicha afirmación?

★5. Una muestra aleatoria de 16 amas de casa indica que 11 prefieren la marca A y 5