5. Introducción a la inferencia estadística

JERZY NEYMAN

Los métodos más utilizados en inferencia estadística son los intervalos de confianza y las pruebas de significación. Ambos métodos son un producto del siglo XX. A partir de un complejo y a veces confuso origen, las pruebas estadísticas tomaron su forma actual en los escritos de R. A. Fisher, al cual nos encontramos al comienzo del capítulo 3. Los intervalos de confianza aparecieron en 1934 gracias al ingenio de Jerzy Neyman (1894-1981).

Neyman se formó en Polonia y, al igual que Fisher, trabajó en un instituto de investigación agrícola. En 1934, Neyman se trasladó a Londres y en 1938 obtuvo una plaza de profesor en la University of California en Berkeley. En EE UU Neyman fundó el Laboratorio de Estadística de Berkeley (*Berkeley's Statistical Laboratory*), del que fue director incluso después de su jubilación en 1961. Ésta no significó una disminución de su actividad científica –permaneció activo hasta el final de su larga vida, e incluso después de jubilado casi llegó a duplicar el número de sus publicaciones. Los problemas estadísticos derivados de campos tan diversos como la astronomía, la biología y la climatología atrajeron la atención de Jerzy Neyman.

A Neyman y a Fisher se les considera los fundadores de la estadística aplicada moderna. Aparte de dar a conocer los intervalos de confianza, Neyman contribuyó a la sistematización de la teoría del muestreo y dio un nuevo enfoque a las pruebas de significación. Fisher, a quien le encantaba la polémica, mostró su desagrado por el enfoque de Neyman, el cual, no siendo tímido, respondió de manera enérgica.

Las pruebas de significación y los intervalos de confianza son los temas de este capítulo. Como la mayoría de los usuarios de la estadística, utilizaremos el método de Fisher para las pruebas de significación. Encontrarás algunas de las ideas de Neyman en la última sección, que es optativa.

5.1 Introducción

Inferir significa sacar conclusiones. La inferencia estadística nos proporciona métodos para sacar conclusiones a partir de datos. Por supuesto que ya hemos estado sacando conclusiones a partir de datos. Lo que es nuevo de la inferencia es que utilizamos la probabilidad para expresar la fuerza de nuestras conclusiones. La probabilidad nos permite tener en cuenta la variación debida al azar y así corregir nuestras conclusiones de acuerdo con los cálculos. He aquí dos ejemplos de cómo la probabilidad puede corregir nuestras conclusiones.

EJEMPLO 5.1

Durante los años de la Guerra de Vietnam un sorteo determinaba el número de orden de incorporación a filas de los hombres. El sorteo asignaba este número escogiendo al azar las fechas de nacimiento. Era de esperar que la correlación entre las fechas de nacimiento y el número de orden de incorporación a filas fuera aproximadamente cero, si el número de orden de incorporación a filas se había escogido al azar. Pero la verdadera correlación entre la fecha de nacimiento y el número de orden de incorporación a filas en el primer sorteo fue $r = -0{,}226$. Es decir, los hombres que habían nacido más tarde dentro de cada año tendían a tener números de orden de incorporación a filas más bajos. Esta pequeña correlación, ¿constituye una evidencia a favor de que el sorteo estaba sesgado? Es difícil saberlo, ya que siempre existe, en la práctica, una pequeña asociación, debida al azar, entre dos variables cualesquiera. Por tanto, calculamos la probabilidad de una correlación como esta en un sorteo al azar. El resultado es menor que 0,001. Puesto que una correlación tan fuerte como la observada no ocurrirá casi nunca en un sorteo al azar, hay una fuerte evidencia de que el sorteo no fue limpio.

Los cálculos de probabilidad también nos permiten evitar conclusiones precipitadas cuando sólo el azar está presente. Se investigaron 20 empresas creadas por mujeres y 20 empresas creadas por hombres. Después de dos años, 12 de las empresas dirigidas por mujeres fracasaron y sólo 8 de las empresas lideradas por hombres tuvieron que cerrar. ¿Podemos concluir a partir de estos datos que las empresas creadas por mujeres suelen quebrar más que las creadas por hombres? Una diferencia de esta magnitud o mayor entre los resultados de estos dos grupos de 20 empresas puede ocurrir una vez de cada cinco por el mero efecto del azar. Un efecto que puede ser fácilmente fruto del azar no es convincente. ■

En este capítulo encontraremos los dos tipos más corrientes de inferencia estadística. La sección 5.2 hace referencia a los *intervalos de confianza* para estimar el valor de un parámetro poblacional. La sección 5.3 presenta las *pruebas de significación*, que valoran la evidencia a favor de una determinada afirmación sobre una población. Los dos tipos de inferencia se basan en las distribuciones de estadísticos. Es decir, los dos tipos de inferencia dan las probabilidades que establecen *lo que ocurriría si utilizáramos el método de inferencia muchas veces.* Este tipo de afirmaciones en términos de probabilidad es característico de la inferencia estadística. Resulta, por ello, necesario comprender el significado de las expresiones probabilísticas que aparecen, por ejemplo, en los resultados de los programas estadísticos.

Los métodos de inferencia necesitan el comportamiento regular a largo plazo que describe la probabilidad. La inferencia es más fiable cuando los datos se obtienen a partir de un diseño aleatorizado adecuado. *Cuando utilizas la inferencia estadística estás actuando como si los datos fueran una muestra aleatoria o procedieran de un experimento aleatorizado.* Si esto no es cierto, tus conclusiones pueden estar expuestas a cualquier tipo de objeción. No te dejes impresionar demasiado por los detalles complejos de la inferencia. Toda esta complicada maquinaria no puede remediar las imperfecciones de la obtención de datos, como las muestras de voluntarios o los diseños sin grupo de control. Utiliza el sentido común que has desarrollado durante el estudio de los tres primeros capítulos de este libro, y empieza a hacer inferencia sólo cuando estés convencido de que los datos se merecen este tipo de análisis.

El objetivo de este capítulo es describir los razonamientos utilizados en la inferencia estadística. Ilustraremos dichos razonamientos con la utilización de algunas técnicas concretas de inferencia. Sin embargo, simplificaremos tanto estas técnicas que no resultarán muy útiles en la práctica. Será en capítulos posteriores cuando mostraremos cómo modificarlas para que resulten útiles en la resolución de casos prácticos. Entonces introduciremos unos métodos de inferencia que pueden ser utilizados en la mayoría de las situaciones con las que nos tropezamos cuando estudiábamos cómo examinar datos. Hay muchas bibliotecas repletas de libros y de manuales de programas estadísticos en los que se presentan gran cantidad de métodos de inferencia mucho más complejos que los que veremos en este libro. Antes de utilizar cualquiera de estos métodos es necesario que comprendas los razonamientos en los que se basan. Los ordenadores harán los cálculos, pero tus conclusiones seguirán dependiendo de tu comprensión de los hechos.

5.2 Estimación con confianza

Los jóvenes tienen más posibilidades de encontrar un trabajo bien pagado si tienen facilidad con los números. ¿Qué habilidades aritméticas tienen los jóvenes americanos en edad de trabajar? Una fuente de datos es la encuesta NAEP (*National Assessment of Educational Progress*), que se hace en EE UU para determinar el nivel educativo de la población. Esta encuesta se basa en una muestra probabilística de hogares a nivel nacional.

EJEMPLO 5.2

La encuesta NAEP incluye una prueba breve de habilidad aritmética y de su aplicación a problemas reales. Los resultados de la prueba van de 0 a 500. Por ejemplo, una persona que obtenga una puntuación de 233 es capaz de sumar los importes de dos cheques que aparecen en el justificante de un banco; alguien que obtenga una puntuación de 325 es capaz de calcular el precio de una comida a partir de los precios de la carta; una persona con una puntuación de 375 puede transformar un precio expresado en centavos por onza a dólares por libra.

En un año reciente participaron en la encuesta NAEP 840 hombres entre 21 y 25 años. La puntuación media de estos hombres en la prueba de cálculo aritmético fue $\bar{x} = 272$. Estos 840 hombres son una muestra aleatoria simple de la población de hombres jóvenes. En base a esta muestra, ¿qué podemos decir sobre la puntuación media μ de la población de los 9.5 millones de hombres jóvenes con estas edades?[1] ■

La media muestral $\bar{x}$ es un estimador insesgado de la media poblacional desconocida μ. Debido a que $\bar{x} = 272$, podemos suponer que μ "está cerca de 272". Para hacer más preciso "cerca de 272", nos preguntamos: *¿Cómo variaría la media muestral $\bar{x}$ si tomáramos muchas muestras de 840 hombres jóvenes de esta misma población?* Recuerda las principales características de la distribución de $\bar{x}$:

- $\bar{x}$ tiene una distribución normal (el teorema del límite central nos indica que el promedio de 840 observaciones tiene una distribución que se parece mucho a una distribución normal).
- La media de esta distribución normal es la misma que la media desconocida de la población μ.
- La desviación típica de $\bar{x}$ en una muestra aleatoria simple de 840 hombres es $\sigma / \sqrt{840}$, donde σ es la desviación típica de las puntuaciones individuales de todos los hombres jóvenes.

Supongamos que sabemos por experiencia que la desviación típica de las puntuaciones de la población de todos los hombres jóvenes es $\sigma = 60$. La desviación típica de $\bar{x}$ es entonces:

[1] Francisco L. Rivera-Batiz, 1992, "Quantitative literacy and the likelihood of employment among young adults", *Journal of Human Resources*, 27, págs. 313-328.

$$\frac{\sigma}{\sqrt{n}} = \frac{60}{\sqrt{840}} = 2.1$$

(No es muy realista suponer que conocemos σ. En el próximo capítulo veremos cómo proceder cuando σ es desconocida. De momento, estamos más interesados en el razonamiento estadístico que en los detalles de los métodos prácticos).

En muchas muestras repetidas de tamaño 840, la puntuación media $\bar{x}$ variaría de acuerdo con una distribución normal de media igual a la media desconocida μ y desviación típica 2.1. La inferencia sobre la μ desconocida utiliza esta distribución de $\bar{x}$. La figura 5.1 presenta esta distribución. Los distintos valores de $\bar{x}$ aparecen a lo largo del eje de las abcisas de la figura y la curva normal indica la probabilidad de estos valores.

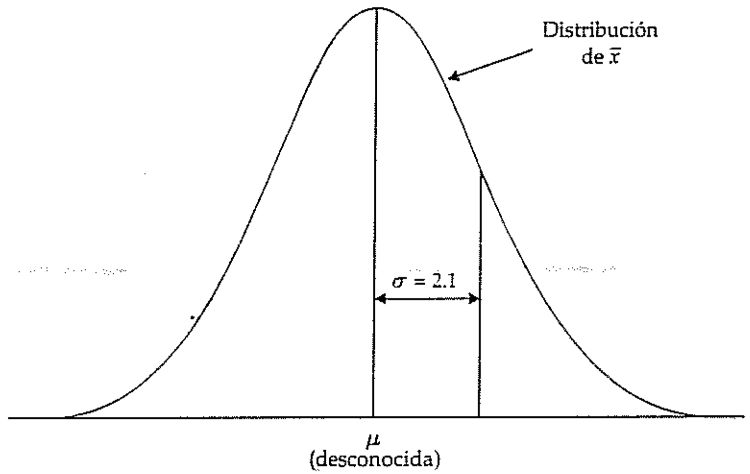


Figura 5.1. Distribución muestral de la puntuación media $\bar{x}$ de una muestra aleatoria simple de 840 hombres jóvenes en la prueba de aritmética de la encuesta NAEP.

5.2.1 Confianza estadística

La figura 5.2 es otra representación de la misma distribución muestral, que ilustra las siguientes ideas:

- La regla del 68-95-99,7 establece que aproximadamente en un 95% de las muestras $\bar{x}$ se encontrará entre $\mu - 2\sigma$ y $\mu + 2\sigma$. Es decir, en un 95% de las muestras $\bar{x}$ estará entre $\mu - 4{,}2$ y $\mu + 4{,}2$.

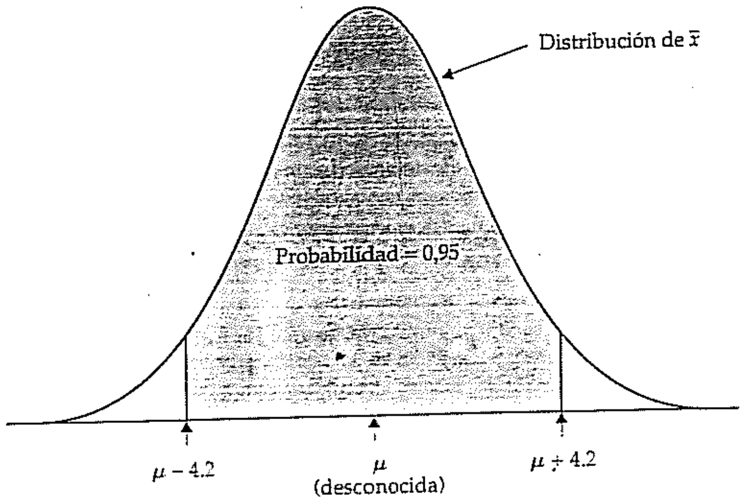


Figura 5.2. En un 95% de las muestras, $\bar{x}$ se encuentra dentro del intervalo $\mu \pm 4.2$. Por tanto, μ se encuentra, también, dentro del intervalo $\bar{x} \pm 4.2$ en estas muestras.

- Siempre que $\bar{x}$ esté situada a una distancia de μ inferior a 4.2, entonces ocurrirá, naturalmente, que μ estará a una distancia de $\bar{x}$ inferior a 4.2. Esto ocurrirá en un 95% de todas las muestras.
- Por tanto, en un 95% de las muestras, la μ desconocida está entre $\bar{x} - 4.2$ y $\bar{x} + 4.2$.

Esta conclusión tan sólo expresa de otra manera una característica de la distribución de $\bar{x}$. El lenguaje de la inferencia estadística utiliza esta característica, que se refiere a lo que ocurriría después de muchas repeticiones, para expresar nuestra confianza en los resultados de cualquier muestra. Nuestra muestra dio $\bar{x} = 272$. Decimos que tenemos una *confianza del 95%* de que la media desconocida de la prueba de aritmética de la encuesta NAEP se encuentre entre:

$$\bar{x} - 4{,}2 = 272 - 4{,}2 = 267{,}8$$

y

$$\bar{x} + 4{,}2 = 272 + 4{,}2 = 276{,}2$$

Asegúrate de que comprendes en qué se basa nuestra confianza. Sólo existen dos posibilidades:

1. El intervalo (267,8, 276,2) contiene la verdadera μ.
2. Nuestra muestra aleatoria simple fue una de las pocas muestras para las cuales $\bar{x}$ no se encuentra a una distancia de μ menor que 4.2. Sólo un 5% de todas las muestras dan resultados tan poco exactos.

No podemos saber si nuestra muestra es una de las 95 muestras de cada 100 para las cuales el intervalo $\bar{x} \pm 4.2$ contiene μ o, en cambio, es una de las muestras desafortunadas que componen el 5% restante. La afirmación de que tenemos una confianza del 95% de que la μ desconocida se encuentre entre 267,8 y 276,2, es una manera breve de decir, "hemos obtenido estos números a partir de un método que funciona correctamente en un 95% de los casos".

Intervalo de confianza

El conjunto de números situados entre los valores $\bar{x} \pm 4.2$ se llama ***intervalo de confianza del 95%*** para μ. Como la mayoría de los intervalos que veremos, éste tiene la estructura

$$\text{estimación} \pm \text{error de estimación}$$

Error de estimación

La estimación ($\bar{x}$ en este caso) es el valor que le suponemos al parámetro desconocido. El ***error de estimación*** ± 4.2 indica la precisión que creemos que tiene nuestra suposición, basada en la variabilidad de la estimación. Éste es un intervalo de confianza del 95% porque contendrá la μ desconocida en un 95% de todas las muestras posibles.

La figura 5.3 ilustra el comportamiento de los intervalos de confianza del 95% en un muestreo repetido. El centro de cada intervalo es $\bar{x}$ y, por tanto, varía de una muestra a otra. La distribución de $\bar{x}$ corresponde a la parte superior de la figura y nos enseña el aspecto de dicha variación después de muchas repeticiones. Los intervalos de confianza del 95% para μ, $\bar{x} \pm 4{,}2$, de 25 muestras aleatorias simples aparecen debajo. El centro $\bar{x}$ de cada intervalo se ha señalado con un punto. Las flechas a ambos lados del punto indican la amplitud del intervalo. De los 25 intervalos de confianza calculados, sólo uno no contiene μ. Con un número muy grande de muestras veríamos que el 95% de los intervalos de confianza contendría μ.

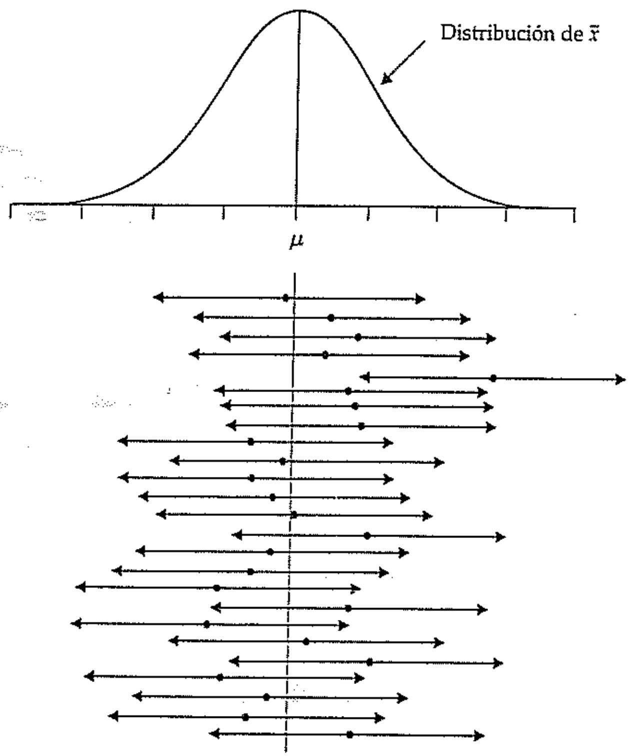


Figura 5.3. Veinticinco muestras de la misma población dieron estos intervalos de confianza del 95%. Después de muchos muestreos, un 95% de las muestras dan intervalos que contienen la media poblacional μ.

EJERCICIOS

5.1. Una encuesta del *New York Times* sobre temas de interés para la mujer entrevistó a 1.025 mujeres seleccionadas aleatoriamente en EE UU, excluyendo Alaska y Hawai. La encuesta halló que el 47% de las mujeres decían que no tenían suficiente tiempo para ellas.

(a) La encuesta daba en sus conclusiones un error de estimación de $\pm$ 3 puntos porcentuales con una confianza del 95%. ¿Cuál es el intervalo de confianza del 95% para el porcentaje de las mujeres adultas que creen que no tienen suficiente tiempo para ellas?

(b) Explícale a alguien que no sepa nada de estadística por qué no podemos decir simplemente que el 47% de las mujeres adultas no tienen suficiente tiempo para ellas.

(c) Luego explica claramente qué quiere decir "una confianza del 95%".

5.2. Un estudiante lee que un intervalo de confianza del 95% para la media de los resultados en la prueba de aritmética de la encuesta NAEP, para hombres entre 21 y 25 años, está entre 267,8 y 276,2. Al preguntarle el significado de este intervalo, el estudiante respondé, "el 95% de todos los hombres jóvenes tienen resultados entre 267,8 y 276,2". ¿Tiene razón el estudiante? Justifica tu respuesta.

5.3. Supón que haces pasar la prueba de aritmética de la encuesta NAEP a una muestra aleatoria simple de 1.000 personas de una gran población, y obtienes una media de 280 y una desviación típica σ = 60. La media $\bar{x}$ de los 1.000 resultados variará si tomas muestras repetidas.

(a) La distribución de $\bar{x}$ es aproximadamente normal. ¿Qué valores toman su media y su desviación típica?

(b) Dibuja la curva normal que describe cómo varía $\bar{x}$ en muchas muestras de esta población. Señala su media y los valores situados a una, dos y tres desviaciones típicas a cada lado de la media.

(c) Según la regla del 68-95-99,7, aproximadamente el 95% de todos los valores de $\bar{x}$ se sitúan entre _______ de la media de esta curva. ¿Cuál es el número que falta? Llama m al error de estimación. Señala la zona entre la media menos m y la media más m en el eje de las abcisas de tu gráfico con una línea gruesa como en la figura 5.2.

(d) Siempre que $\bar{x}$ se sitúe en la zona que has señalado, el verdadero valor de la media de la población, μ = 280, se hallará en el intervalo de confianza $\bar{x} - m$ y $\bar{x} + m$. Debajo de tu gráfico, dibuja el intervalo de confianza de un valor de $\bar{x}$ que esté situado dentro de la zona señalada en (c) y de un valor de $\bar{x}$ que esté situado fuera (utiliza la figura 5.3 como modelo).

(e) ¿En qué porcentaje de todas las muestras el intervalo de confianza $\bar{x} \pm m$ contendrá a la verdadera media μ = 280?

5.4. Los óxidos de nitrógeno (denominados NO_X de forma abreviada) que emiten los automóviles y los camiones contribuyen de manera importante a la contaminación del aire. La cantidad de NO_X que emite un determinado modelo de automóvil varía entre los distintos vehículos. Para un modelo de camión ligero, las emisiones de NO_X varían con una media μ que es desconocida y una desviación típica σ = 0,4 gramos por km. En una muestra aleatoria simple de 50 de estos camiones, la media muestral $\bar{x}$ del nivel de NO_X representa una estimación de la μ desconocida. Obtendrás distintos valores de $\bar{x}$ si repites tu muestreo.

(a) La distribución de $\bar{x}$ es aproximadamente normal. ¿Cuáles son su media y su desviación típica?

(b) Dibuja la curva normal de la distribución de $\bar{x}$. Señala su media y los valores situados a una, dos y tres desviaciones típicas a cada lado de la media.

(c) Según la regla del 68-95-99,7, aproximadamente el 95% de todos los valores de $\bar{x}$ se hallan a menos de una distancia m de la media de la distribución muestral. ¿Cuál es el valor de m? Señala los puntos del eje de las abcisas de tu gráfico que están a una distancia de la media menor que m, como en la figura 5.2.

(d) Siempre que $\bar{x}$ se sitúe en la zona que has señalado, la media desconocida de la población μ se encontrará en el intervalo de confianza $\bar{x} \pm m$. ¿Para qué porcentaje de las muestras ocurre esto?

(e) Siguiendo el estilo de la figura 5.3, dibuja los intervalos de confianza debajo de tu gráfico para dos valores de $\bar{x}$, uno que esté situado dentro de la zona señalada y uno que esté situado fuera de ella.

5.2.2 Intervalos de confianza

Nivel de confianza

En estadística se han construido intervalos de confianza, para muchos parámetros distintos, basados en diversas formas de obtener datos. Hablaremos de algunos de ellos en los próximos capítulos. Cualquier intervalo de confianza tiene dos partes: un *intervalo* calculado a partir de los datos y un ***nivel de confianza*** que da la probabilidad de que el método produzca un intervalo que contenga el parámetro. Al utilizar estos intervalos se escoge el nivel de confianza, que muy a menudo es del 90% o superior, ya que queremos estar suficientemente seguros de nuestras conclusiones. Llamaremos C al nivel de confianza expresado en tanto por uno. Por ejemplo, a un intervalo de confianza del 95% le corresponde un valor C igual a 0,95. He aquí la definición de un intervalo de confianza para un parámetro poblacional desconocido. En nuestros ejemplos, este parámetro es la media poblacional μ, pero también podría ser una proporción poblacional p, la desviación típica de una población σ, o cualquier otro parámetro.

INTERVALO DE CONFIANZA

Un intervalo de confianza de nivel C para un parámetro poblacional es un intervalo calculado a partir de los datos de una muestra por un método que tiene una probabilidad C de producir un intervalo que contenga el verdadero valor del parámetro.

Ahora podemos dar la fórmula para calcular un intervalo de confianza de nivel C para la media poblacional μ de una población, cuando los datos son una muestra aleatoria simple de tamaño n. El intervalo se basa en el hecho de que la distribución de la media muestral $\bar{x}$ es aproximadamente normal. Para obtener el nivel de confianza C debemos tomar los elementos de la población que se encuentran en la parte central de probabilidad C de una distribución normal. Para conseguirlo, tenemos que desplazarnos z^* desviaciones típicas a ambos lados de la media. El número z^* es el mismo para cualquier distribución normal, de manera que podemos utilizar la tabla de la distribución normal estandarizada. He aquí un ejemplo de cómo hallar z^*.

EJEMPLO 5.3

Para hallar un intervalo de confianza del 80% debemos tener el 80% central de los valores de la distribución normal de $\bar{x}$. Si tenemos el 80% central, dejamos fuera un 20%; un 10% en cada cola de la distribución. Por tanto, z^* es el punto que deja un área de 0.1 a su derecha (y un área a su izquierda de 0.9) por debajo de una curva normal estandarizada. En el cuerpo central de la tabla A hallarás el punto que deja un área a su izquierda igual a 0,9. El valor más próximo es $z^* = 1{,}28$. Hay un área de 0,8 por debajo de la curva normal estandarizada entre − 1,28 y 1,28. La figura 5.4 ilustra la relación entre z^* y las áreas delimitadas por debajo de la curva. ■

La figura 5.5 ilustra el caso general para cualquier nivel de confianza C. Si capturamos el área central C, el área de las colas es de $1 - C$, o de $(1 - C)/2$ en cada cola. Para cada valor de C puedes hallar los valores de z^* en la tabla A. He aquí los resultados para los niveles de confianza más frecuentes:

Nivel de confianza	Área de la cola	z^*
90%	0,05	1,645
95%	0,025	1,960
99%	0,005	2,576

Fíjate en que para una confianza del 95% utilizamos $z^* = 1{,}960$. Esto es más exacto que el valor aproximado $z^* = 2$ dado por la regla del 68-95-99,7. La última fila de la tabla C da los valores de z^* para muchos niveles de confianza C. Esta fila está encabezada por z^* (encontrarás la tabla C al final del libro. Utilizaremos las otras filas de la tabla en el próximo capítulo). A pesar de que podemos hallar z^* en la tabla C, justo encima de los valores de los niveles de confianza C, es habitual describir el punto z^* en términos del área situada a su derecha. Por ejemplo, llamamos a 1,960 el *valor crítico superior* de 0,025 de la distribución normal estandarizada.

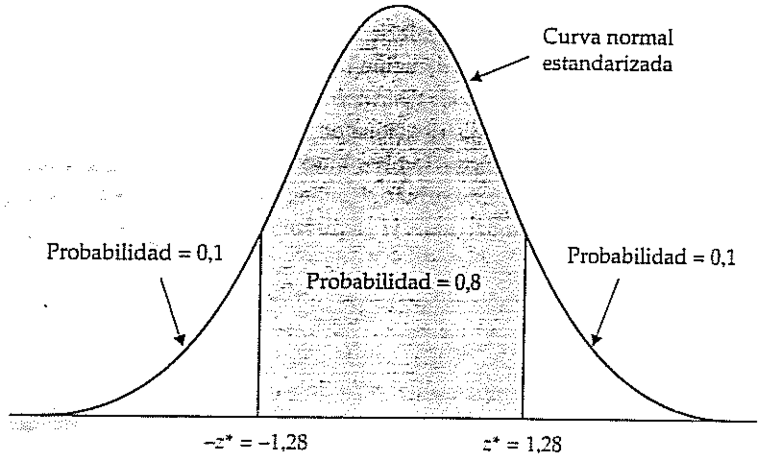


Figura 5.4. El 0,8 de la probabilidad central de una curva normal estandarizada se encuentra entre −1,28 y 1,28. A la derecha de 1,28 y por debajo de la curva hay un área de 0,1.

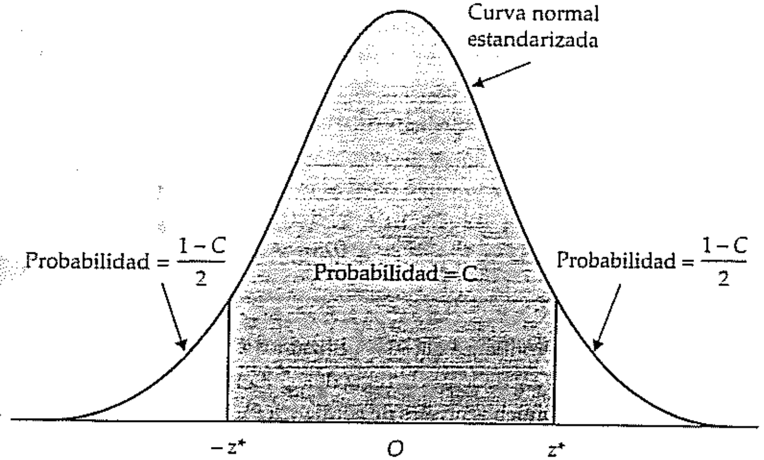
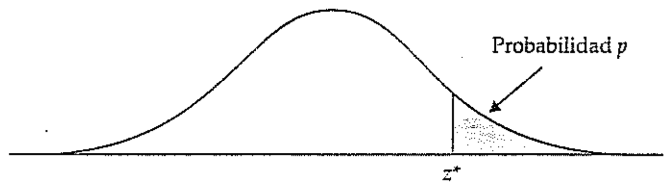


Figura 5.5. En general, la probabilidad central C, por debajo de una curva normal estandarizada, se encuentra entre $-z^*$ y z^*. Debido a que z^* tiene un área de $(1-C)/2$ a su derecha por debajo de la curva, llamamos a z^* el valor crítico superior.

VALORES CRÍTICOS

El número z^*, con una probabilidad p a su derecha y por debajo de la curva normal estandarizada, se llama el **valor crítico superior de** p de la distribución normal estandarizada.



He aquí cómo calcular un intervalo de confianza de nivel C.

- Cualquier curva normal tiene una probabilidad C entre el punto situado a z^* desviaciones típicas a la izquierda de su media y el punto situado a z^* desviaciones típicas a la derecha de su media.
- La desviación típica de la distribución de $\bar{x}$ es $\sigma/\sqrt{n}$, y su media es la media de la población μ. Por tanto, existe una probabilidad C de que la media muestral observada $\bar{x}$ tome un valor entre

$$\mu - z^* \frac{\sigma}{\sqrt{n}} \quad \text{y} \quad \mu + z^* \frac{\sigma}{\sqrt{n}}$$

- Siempre que ocurre lo anterior, la media poblacional μ se encuentra entre

$$\bar{x} - z^* \frac{\sigma}{\sqrt{n}} \quad \text{y} \quad \bar{x} + z^* \frac{\sigma}{\sqrt{n}}$$

Éste es nuestro intervalo de confianza. La estimación de la media desconocida μ es $\bar{x}$, y el error de estimación es $z^*\sigma/\sqrt{n}$.

INTERVALO DE CONFIANZA PARA LA MEDIA POBLACIONAL

Obtén una muestra aleatoria simple de tamaño n de una población con una media desconocida μ y una desviación típica σ conocida. Un intervalo de confianza de nivel C para μ es:

$$\bar{x} \pm z^* \frac{\sigma}{\sqrt{n}}$$

Aquí z^* es el valor crítico superior de $(1-C)/2$ de la distribución normal estandarizada, hallado en la tabla C. Este intervalo es exacto cuando la distribución poblacional es normal y aproximadamente correcto para n grande en los demás casos.

EJEMPLO 5.4

Un fabricante de productos farmacéuticos analiza un comprimido de cada uno de los lotes de un medicamento, para verificar la concentración de la materia activa de los comprimidos. El método de análisis químico no es totalmente preciso. Los análisis repetidos de un mismo comprimido dan resultados ligeramente distintos. Los resultados de análisis repetidos de un mismo comprimido siguen aproximadamente una distribución normal. El método de análisis no tiene sesgo; por tanto, la media μ de la población de todos los análisis es la verdadera concentración de materia activa en un comprimido. Se conoce que la desviación típica de la distribución de los análisis es $\sigma = 0{,}0068$ gramos por litro. En la rutina del laboratorio cada comprimido se analiza 3 veces y se calcula su media.

Tres análisis de un comprimido dan las concentraciones

0.8403 0.8363 0.8447

Queremos un intervalo de confianza del 99% para la verdadera concentración μ.

La media muestral de los tres análisis es:

$$\bar{x} = \frac{0.8403 + 0.8363 + 0.8447}{3} = 0.8404$$

Para un intervalo de confianza del 99%, vemos en la tabla C que $z^* = 2{,}576$. Por tanto, un intervalo de confianza del 99% para μ es:

$$\bar{x} \pm z^* \frac{\sigma}{\sqrt{n}} = 0{,}8404 \pm 2{,}576 \, \frac{0{,}0068}{\sqrt{3}} =$$
$$= 0{,}8404 \pm 0{,}0101 =$$
$$= (0{,}8303,\ 0{,}8505)$$

Tenemos una confianza del 99% de que el verdadero valor de la concentración de materia activa se halla entre 0,8303 y 0,8505 gramos por litro. ■

Supón que el resultado de un solo análisis diera $x = 0{,}8404$, el mismo valor que la media calculada en el ejemplo 5.4. Repitiendo el cálculo anterior pero para $n = 1$, obtenemos que el intervalo de confianza del 99% basado en un único análisis es:

$$\bar{x} \pm z^* \frac{\sigma}{\sqrt{n}} = 0{,}8404 \pm 2{,}576 \, \frac{0{,}0068}{\sqrt{1}} =$$
$$= 0{,}8404 \pm 0.0175 =$$
$$= (0{,}8229,\ 0{,}8579)$$

La media de tres lecturas da un error de estimación menor y, por tanto, un intervalo de confianza más corto que el de una sola lectura. La figura 5.6 ilustra la ganancia de precisión cuando se utilizan tres observaciones.

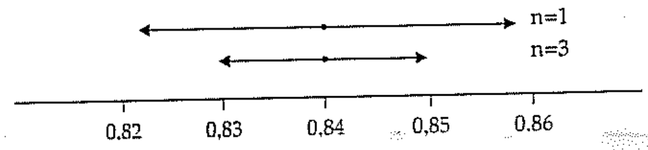


Figura 5.6. Los intervalos de confianza para $n = 1$ y $n = 3$ del ejemplo 5.4. Las muestras mayores dan intervalos de confianza más cortos.

La forma de los intervalos de confianza para la media poblacional μ se basa en el hecho de que el estadístico $\bar{x}$ utilizado para estimar μ tiene una distribución normal. Debido a que muchos estadísticos muestrales tienen distribuciones normales (al menos aproximadamente), es útil observar que el intervalo de confianza tiene la forma:

$$\text{estimación} \pm z^* \sigma_{\text{de la estimación}}$$

La estimación basada en la muestra es el centro del intervalo de confianza. El error de estimación es $z^* \sigma_{\text{de la estimación}}$. El nivel de confianza deseado C determina el valor z^* de la tabla C. La desviación típica de la estimación, $\sigma_{\text{de la estimación}}$, depende de la estimación concreta que utilicemos. Cuando la estimación de una muestra aleatoria simple es $\bar{x}$, la desviación típica de la estimación es $\sigma/\sqrt{n}$.

EJERCICIOS

5.5. Un estudio sobre la carrera profesional de los directores de hotel envió cuestionarios a una muestra aleatoria simple de 160 hoteles pertenecientes a las principales cadenas hoteleras de EE UU. Hubo 114 respuestas. El promedio de tiempo que estos 114 directores habían pasado en su empresa actual era de 11,78 años. Da un intervalo de confianza del 99% para el número medio de años que los directores de hotel de las principales cadenas han estado en su empresa actual (supón que se sabe que la desviación típica del tiempo de permanencia de los directores en la empresa es de 3,2 años).

5.6. La DRP (*Degree of Reading Power*) es una prueba sobre la capacidad lectora de los niños. Aquí tienes los resultados de la prueba DRP de una muestra de 44 alumnos de tercer curso de un mismo municipio.[2]

40	26	39	14	42	18	25	43	46	27	19
47	19	26	35	34	15	44	40	38	31	46
52	25	35	35	33	29	34	41	49	28	52
47	35	48	22	33	41	51	27	14	54	45

(a) Creemos que la distribución de los resultados en la prueba DRP es aproximadamente normal. Dibuja un diagrama de tallos o un histograma de la distribución de estos 44 resultados y describe su forma.

(b) Supón que se sabe que la desviación típica de la población de los resultados de la prueba DRP es $\sigma = 11$. Da un intervalo de confianza del 99% para la media de los resultados en el municipio.

(c) ¿Confiarías en tus conclusiones de (b) si estos resultados provinieran de una misma clase en una escuela del municipio? ¿Por qué?

5.7. Aquí tienes las medidas (en milímetros) de una dimensión crítica de una muestra de cigüeñales de automóvil.

224,120	224,001	224,017	223,982	223,989	223,961
223,960	224,089	223,987	223,976	223,902	223,980
224,098	224,057	223,913	223,999		

Los datos provienen de un proceso de producción que se sabe que tiene una desviación típica $\sigma = 0{,}060$ mm. La media del proceso se supone que es $\mu = 224$ mm, pero puede desviarse de su objetivo durante la producción.

(a) Suponemos que la distribución de la dimensión crítica de los cigüeñales es aproximadamente normal. Dibuja un diagrama de tallos o un histograma con estos datos y describe la forma de la distribución.

(b) Da un intervalo de confianza del 95% para la media del proceso en el momento en que se produjeron estos cigüeñales.

[2] Maribeth Cassidy Schmitt, *The Effects of an Elaborated Directed Reading Activity on the Metacomprehension Skills of Third Graders*, Ph.D., Purdue University, 1987.

5.8. El análisis del nivel de potasio en la sangre no es absolutamente preciso. Además, el nivel de potasio en la sangre de una persona varía ligeramente de un día para otro. Supón que el nivel de potasio en la sangre de una misma persona en análisis repetidos en distintos días varía de forma normal con $\sigma = 0{,}2$.

(a) Se analiza una vez el nivel de potasio de Julia. El resultado es $x = 3{,}2$. Da un intervalo de confianza del 90% para la media de su nivel de potasio.

(b) Si se hubieran hecho 3 análisis en días distintos y la media de los análisis fuera $\bar{x} = 3{,}2$, ¿cuál es el intervalo de confianza del 90% para la media del nivel de potasio en la sangre de Julia?

5.2.3 Comportamiento de los intervalos de confianza

El intervalo de confianza $\bar{x} \pm z^*\sigma/\sqrt{n}$ para la media de una población normal ilustra algunas de las propiedades importantes que son compartidas por todos los intervalos de confianza de uso frecuente. El usuario escoge el nivel de confianza, y el error de estimación depende de esta decisión. Nos gustaría tener un nivel de confianza alto y también un error de estimación pequeño. Un nivel de confianza alto significa que nuestro método casi siempre da respuestas correctas. Un error de estimación pequeño significa que la estimación del parámetro poblacional es bastante precisa. El error de estimación es

$$\text{Error de estimación} = z^* \frac{\sigma}{\sqrt{n}}$$

Esta expresión tiene z^* y σ en el numerador y $\sqrt{n}$ en el denominador. Por tanto, el error de estimación se hace menor cuando

- z^* se hace menor. Una z^* menor es lo mismo que un nivel de confianza C menor (mira otra vez la figura 5.5). Existe una relación entre el nivel de confianza y el error de estimación. Con unos mismos datos, para tener un error de estimación menor, tienes que aceptar una confianza menor.
- σ se hace menor. La desviación típica σ mide la variación de la población. Puedes pensar en la variación entre los individuos de una población como en un ruido que oculta el valor promedio μ. Es más fácil estimar con precisión μ cuando σ es pequeña.
- n se hace mayor. Un incremento del tamaño de la muestra n reduce el error de estimación para un nivel de confianza determinado. Debido a que n está dentro de la raíz cuadrada, tenemos que multiplicar por cuatro el tamaño de la muestra para reducir a la mitad el error de estimación.

EJEMPLO 5.5

Supón que el fabricante de productos farmacéuticos del ejemplo 5.4 considera que un nivel de confianza del 90%, en vez de un nivel del 99%, ya es suficiente. La tabla C da el valor crítico para una confianza del 90%, que es $z^* = 1{,}645$. El intervalo de confianza del 90% para μ, basado en tres análisis repetidos con una media $\bar{x} = 0.8404$, es

$$\begin{aligned}\bar{x} \pm z^*\sigma/\sqrt{n} &= 0{,}8404 \pm (1{,}645)(0{,}0068/\sqrt{3}) = \\ &= 0{,}8404 \pm 0{,}0065 = \\ &= (0{,}8339,\ 0{,}8469)\end{aligned}$$

Al pasar de una confianza del 99% a una del 90%, el error de estimación se ha reducido de ± 0.0101 a $\pm 0{,}0065$. La figura 5.7 compara estos dos intervalos.

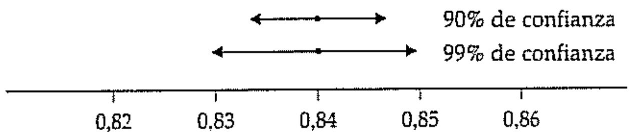


Figura 5.7. Los intervalos de confianza del 90% y del 99% del ejemplo 5.5. Una mayor confianza requiere una mayor amplitud del intervalo.

Aumentar el número de observaciones de 3 a 12 reduce, también, la amplitud del intervalo de confianza del 99% del ejercicio 5.4. Comprueba que sustituyendo $\sqrt{3}$ por $\sqrt{12}$, el error de estimación $\pm 0{,}0101$ se reduce a la mitad, debido a que ahora tenemos cuatro veces más observaciones. ■

EJERCICIOS

5.9. Los ejemplos 5.4 y 5.5 dan intervalos de confianza para la concentración en materia activa de unos comprimidos μ, basados en 3 lecturas con $\bar{x} = 0{,}8404$ y $\sigma = 0{,}0068$. El intervalo de confianza del 99% va de 0,8303 a 0,8505 y el intervalo de confianza del 90% va de 0,8339 a 0,8469.

(a) Halla un intervalo de confianza del 80% para μ.

(b) Halla un intervalo de confianza del 99,9% para μ.

(c) Dibuja un gráfico como el de la figura 5.7 que compare los cuatro intervalos.

¿Cómo afecta el aumento del nivel de confianza a la amplitud del intervalo de confianza?

5.10. Halla el error de estimación para una confianza del 99% con los datos del ejemplo 5.4, si el laboratorio mide la concentración de cada comprimido 12 veces. Comprueba que tu resultado es la mitad del error de estimación basado en los tres análisis del ejemplo 5.4.

5.11. La prueba de aritmética de la encuesta NAEP (ejemplo 5.2) también se hizo pasar a una muestra de 1.077 mujeres entre 21 y 25 años. La media de sus resultados fue 275. Supón que la desviación típica de todos los resultados individuales es $\sigma = 60$.

(a) Da un intervalo de confianza del 95% para la media de los resultados μ de la población de todas las mujeres jóvenes.

(b) Da intervalos de confianza del 90% y del 99% para μ.

(c) ¿Cuáles son los errores de estimación para una confianza del 90%, del 95% y del 99%? ¿Cómo afecta el aumento del nivel de confianza al error de estimación de un intervalo de confianza?

5.12. La muestra NAEP de 1.077 mujeres jóvenes tenía una media en la prueba de aritmética $\bar{x} = 275$. Supón que la desviación típica de todos los resultados individuales es $\sigma = 60$.

(a) Da un intervalo de confianza del 95% para la media de los resultados μ de la población de las mujeres jóvenes.

(b) Supón que el mismo resultado, $\bar{x} = 275$, provenga de una muestra de 250 mujeres. Da el intervalo de confianza del 95% para la media de la población μ en este caso.

(c) Luego, supón que una muestra de 4.000 mujeres hubiera dado la media muestral $\bar{x} = 275$ y da, de nuevo, un intervalo de confianza del 95% para μ.

(d) ¿Cuáles son los errores de estimación para muestras de tamaño 250, 1.077 y 4.000? ¿Cómo afecta el aumento del tamaño de la muestra al error de estimación de un intervalo de confianza?

5.2.4 Elección del tamaño de la muestra

Un usuario prudente de los métodos estadísticos nunca planifica la obtención de los datos sin planear, al mismo tiempo, la inferencia. Si obtienes suficientes observaciones puedes conseguir a la vez un nivel de confianza elevado y un error de estimación pequeño. El error de estimación de un intervalo de confianza para la media de una población que se distribuye normalmente es $m = z^*\sigma/\sqrt{n}$. Para obtener un error de estimación deseado m, sustituye el valor de z^* por el valor que le corresponde según el nivel de confianza escogido y despeja n en la ecuación anterior. He aquí el resultado.

TAMAÑO DE LA MUESTRA PARA UN ERROR DE ESTIMACIÓN DESEADO

Un intervalo de confianza para la media poblacional tendrá un error de estimación determinado m, cuando el tamaño de la muestra sea

$$n = \left(\frac{z^*\sigma}{m}\right)^2$$

Esta fórmula no se puede utilizar a la ligera. En la práctica, la obtención de observaciones cuesta tiempo y dinero. Puede ocurrir que el tamaño de la muestra ideal sea inviable por razones económicas. De nuevo, fíjate en que es el tamaño de la *muestra* lo que determina el error de estimación. El tamaño de la *población* (siempre que la población sea mucho mayor que la muestra) no influye sobre el tamaño de la muestra que necesitamos.

EJEMPLO 5.6

La gerencia de la empresa exige al laboratorio del ejemplo 5.4 que dé resultados con una precisión de $\pm 0{,}005$ y con una confianza del 95%. ¿De cuántas observaciones tienen que constar las muestras?

El error de estimación deseado es $m = 0{,}005$. Para una confianza del 95%, la tabla C da $z^* = 1{,}960$. Sabemos que $\sigma = 0{,}0068$. Por tanto,

$$n = \left(\frac{z^*\sigma}{m}\right)^2 = \left(\frac{1{,}96 \times 0{,}0068}{0{,}005}\right)^2 = 7{,}1$$

Debido a que 7 observaciones darán un error de estimación ligeramente superior al deseado, y 8 observaciones un error de estimación algo menor, el laboratorio tiene que hacer 8 análisis de cada comprimido para cumplir con las exigencias de la gerencia (siempre redondea n hasta el número entero *mayor* más próximo). Cuando la gerencia conozca el coste de realizar tantos análisis, puede ser que reconsidere su demanda. ■

EJERCICIOS

5.13. Para valorar la precisión de una balanza de laboratorio, se pesa repetidamente una pesa con un peso conocido de 10 gramos. Las lecturas de la balanza se distribuyen de forma normal con una media desconocida (esta media es de 10 gramos si la balanza no está sesgada). La desviación típica de las lecturas de la balanza se sabe que es de 0,0002 gramos.

(a) Se pesa 5 veces la pesa. El resultado medio es de 10,0023 gramos. Da un intervalo de confianza del 98% para la media de las pesadas repetidas de la pesa.

(b) ¿Cuántas pesadas deben promediarse para obtener un error de estimación de $\pm 0{,}0001$ con una confianza del 98%?

5.14. ¿Cuál debería ser el tamaño de la muestra de los directores de hotel del ejercicio 5.5, para estimar la media μ con una precisión de ± 1 año con una confianza del 99%?

5.15. ¿Cuál debería ser el tamaño de la muestra de los cigüeñales del ejercicio 5.7, para estimar la media μ con una precisión de $\pm 0{,}020$ mm con una confianza del 95%?

5.2.5 Algunas precauciones

Cualquier fórmula para hacer inferencia estadística es correcta sólo en unas circunstancias concretas. Si los procedimientos estadísticos llevaran advertencias como los medicamentos, la mayoría de los métodos de inferencia llevarían advertencias muy extensas. La fórmula que ya conocemos para la estimación de una media normal, $\bar{x} \pm z^*\sigma/\sqrt{n}$, va acompañada de la siguiente lista de advertencias.

- Los datos deben proceder de una muestra aleatoria simple de una población. Estamos completamente a salvo si los datos proceden de una muestra aleatoria simple. No estamos en gran peligro si los datos se han obtenido de forma que se puedan asimilar a los de una muestra aleatoria de una población. Éste es el caso de los ejemplos 5.4, 5.5 y 5.6, donde teníamos presente que la población era el resultado de un número muy grande de análisis repetidos de un mismo comprimido.
- La fórmula no es correcta para sistemas de muestreo más complejos que los muestreos aleatorios simples, pero existen métodos correctos para esos otros sistemas. No mostraremos cómo calcular intervalos de confianza en el caso de muestreos estratificados o de muestreos en etapas múltiples. Si utilizas estos tipos de muestreo, asegúrate de que sabes cómo llevar a cabo la inferencia.

- A partir de muestras obtenidas de forma caprichosa, sin seguir ningún tipo de diseño estadístico, es imposible hacer inferencia. No hay fórmulas maravillosas que nos permitan utilizar datos obtenidos de forma incorrecta.
- Debido a que las observaciones atípicas tienen una gran influencia sobre el valor de $\bar{x}$, éstas pueden tener un gran efecto sobre los intervalos de confianza. Antes de calcular un intervalo de confianza tienes que averiguar si hay observaciones atípicas. Siempre que sea posible tienes que corregir sus valores o justificar su eliminación. Si no pueden ser eliminadas, consulta con un experto en estadística sobre los métodos que no son sensibles a las observaciones atípicas.
- Si el tamaño de la muestra es pequeño y la población no es normal, el verdadero nivel de confianza del intervalo será distinto del valor C utilizado para calcular el intervalo. Examina cuidadosamente tus datos. Fíjate en la asimetría y en otros indicadores de falta de normalidad. El intervalo se basa, sólo, en la distribución de $\bar{x}$, que incluso para muestras pequeñas es más normal que las observaciones individuales. Cuando $n \geq 15$, el nivel de confianza del intervalo no se ve muy afectado por la falta de normalidad de la población, a no ser que la población sea muy asimétrica, o que existan observaciones atípicas extremas. Este tema lo discutiremos con más detalle en el próximo capítulo.
- Tienes que conocer la desviación típica σ de la población. Esta exigencia poco realista hace que el intervalo $\bar{x} \pm z^*\sigma/\sqrt{n}$ sea de poca utilidad en la práctica estadística. En el próximo capítulo veremos qué hay que hacer cuando σ es desconocida. De todas formas, si la muestra es grande, la desviación típica muestral s estará próxima a la σ desconocida. En estos casos $\bar{x} \pm z^*s/\sqrt{n}$ es un intervalo aproximado para μ.

La precaución más importante, en relación con los intervalos de confianza, es una consecuencia de la primera de estas advertencias. *El error de estimación de un intervalo de confianza sólo tiene en cuenta el error del muestreo aleatorio.* El error de estimación se obtiene a partir de la distribución muestral e indica cuál es la magnitud esperada del error debida a la variabilidad en la obtención aleatoria de los datos. Dificultades prácticas, tales como la no-respuesta o la falta de cobertura de una encuesta, pueden causar errores adicionales que podrían ser mayores que el error del muestreo aleatorio. Recuerda este hecho desagradable cuando leas los resultados de las encuestas de opinión o de otros tipos de encuestas. Los problemas prácticos que se presentan durante la realización de las encuestas pueden influir de forma muy importante en la credibilidad de sus resultados, pero no se tienen en cuenta en el cálculo del error de estimación.

Cualquier procedimiento de inferencia estadística debería ir acompañado de su propia lista de advertencias. Debido a que muchas de las advertencias son similares

a las del listado anterior, cada vez que presentemos un nuevo procedimiento de inferencia no reproduciremos el listado completo. Es fácil establecer (a partir de las matemáticas de la probabilidad) las condiciones bajo las cuales un determinado método de inferencia es exactamente correcto. Estas condiciones *nunca* se cumplen totalmente en la práctica. Por ejemplo, no existe ninguna población que sea exactamente normal. La decisión de cuándo un procedimiento puede ser utilizado en la práctica requiere, a menudo, un cuidadoso análisis exploratorio de los datos.

Finalmente, tienes que comprender perfectamente lo que la confianza estadística no dice. Tenemos una confianza del 95% en que la media de la prueba de aritmética de la encuesta NAEP, para todos los hombres entre 21 y 25 años, se encuentra entre 267,8 y 276.2. Es decir, estos números se calcularon mediante un método que da respuestas correctas en un 95% de todas las muestras posibles. *No podemos* decir que la probabilidad de que la media poblacional μ se encuentre entre 267.8 y 276.2, sea del 95%. Después de haber escogido una determinada muestra y haber calculado el intervalo, desaparece el azar. La verdadera media puede estar o no contenida entre 267,8 y 276.2. Los cálculos de probabilidad en la inferencia estadística describen con qué frecuencia el *método* da una respuesta correcta.

EJERCICIOS

5.16. En un debate radiofónico se invita a los oyentes a que participen en una discusión sobre una propuesta de aumento del sueldo de los concejales de un ayuntamiento. "¿Qué salario anual crees que se debería pagar a los concejales? Llámanos con tu cifra". Llama un total de 958 personas. La media del salario que sugieren las personas que llaman es $\bar{x}$ = 8.740 dólares al año, con una desviación típica de las respuestas s = 1.125 dólares. Para una gran muestra como ésta, s se acerca mucho a la desviación típica desconocida de la población σ. La emisora dice que el intervalo de confianza del 95% para el salario medio μ de los concejales que propondrían todos los ciudadanos, va de 8.699 a 8.811 dólares.

(a) ¿Es correcto el cálculo de la emisora?

(b) ¿Sus conclusiones describen la población de todos los habitantes de la ciudad? Justifica tu respuesta.

5.17. En 1976, las elecciones presidenciales de EEUU, en las que se enfrentaron Jimmy Carter y Gerald Ford, se ganaron sólo por un pequeño margen. Una encuesta realizada inmediatamente antes de estos comicios reveló que el 51% de la muestra tenía la intención de votar a Carter. La empresa encuestadora anunció que tenía una certeza del 95% de que este resultado estaba a menos de ± 2 puntos porcentuales del verdadero porcentaje de votantes a favor de Carter.

(a) Utilizando un lenguaje sencillo, explícale a alguien que no sepa estadística qué significa "una certeza del 95%" en este caso.

(b) La encuesta mostraba que Carter iba en cabeza. Sin embargo, la empresa encuestadora dijo que los resultados eran demasiado ajustados como para predecir quién iba a ganar. Explica por qué.

(c) Al oír los resultados de la encuesta, un político preguntó nervioso: "¿cuál es la probabilidad de que más de la mitad de los votantes prefiera a Carter?" Un experto en estadística contestó que a esa pregunta no se podía responder a partir de los resultados de la encuesta y que no tenía sentido hablar de tal probabilidad. Explica por qué.

5.18. Una encuesta reciente del *New York Times*/CBS planteó la pregunta: "¿estarías a favor de introducir una enmienda en la Constitución que permitiera los rezos organizados en las escuelas públicas?" El 66% de la muestra contestó "Sí". El artículo que describe la encuesta dice que "está basada en entrevistas telefónicas llevadas a cabo entre el 13 y el 18 de septiembre a 1.664 adultos de EE UU, exceptuando Alaska y Hawai... los números telefónicos se construyeron de forma aleatoria y, por tanto, accediendo a números que aparecían o no en la guía telefónica".

(a) El artículo da un error de estimación de 3 puntos porcentuales. Determina un intervalo de confianza para el porcentaje de todos los adultos que están a favor de la enmienda sobre los rezos organizados en las escuelas.

(b) El artículo continúa diciendo: "los errores teóricos no tienen en cuenta un error de estimación adicional resultante de las diversas dificultades prácticas que surgen al llevar a cabo cualquier encuesta de opinión pública". Haz una lista de algunas "dificultades prácticas" que puedan causar errores a añadir al error de estimación del $\pm 3\%$. Presta especial atención a la descripción del método de muestreo que da el artículo.

RESUMEN

El objetivo de un **intervalo de confianza** es estimar un parámetro desconocido obteniendo una indicación sobre la precisión de la estimación y sobre cuál es nuestra confianza de que el resultado sea correcto.

Cualquier intervalo de confianza tiene dos partes: el intervalo calculado a partir de los datos y el nivel de confianza. Los **intervalos**, a menudo, tienen la forma:

$$\text{estimación} \pm \text{error de estimación}$$

El **nivel de confianza** indica la probabilidad de que el método dé una respuesta correcta. Esto es, si utilizaras repetidamente los intervalos de confianza del 95%, después de muchos muestreos, un 95% de estos intervalos contendría el verdadero valor del parámetro. No puedes saber si un intervalo de confianza del 95% calculado a partir de un determinado conjunto de datos contiene el verdadero valor del parámetro.

Un intervalo de confianza de nivel C para la media μ de una población normal con una desviación típica σ conocida, basado en una muestra aleatoria simple de tamaño n, viene dado por

$$\bar{x} \pm z^* \frac{\sigma}{\sqrt{n}}$$

Aquí, z^* se ha escogido de manera que la curva normal estandarizada tenga un área C entre $-z^*$ y z^*. Debido al teorema del límite central, este intervalo es aproximadamente correcto para muestras grandes cuando la población no es normal.

El número z^* se llama **valor crítico superior** de p de la distribución normal estandarizada para $p = (1 - C)/2$. En la tabla C aparecen los valores críticos de muchos niveles de confianza.

Si se mantiene lo demás constante, el **error de estimación** de un intervalo de confianza se hace pequeño cuando

- el nivel de confianza C disminuye,
- la desviación típica poblacional σ disminuye, y
- el tamaño de la muestra n aumenta.

El tamaño de muestra necesario para obtener un intervalo de confianza con un determinado error de estimación m para una media normal es

$$n = \left(\frac{z^* \sigma}{m}\right)^2$$

Donde z^* es el valor crítico para el nivel de confianza deseado. Redondea siempre n hacia arriba cuando utilices esta fórmula.

La fórmula de un determinado intervalo de confianza es correcta sólo en unas condiciones concretas. Las condiciones más importantes hacen referencia al procedimiento utilizado para obtener los datos. Otros factores tales como la forma de la distribución de la población también pueden ser importantes.

EJERCICIOS DE LA SECCIÓN 5.2

5.19. La prueba ARSMA (*Acculturation Rating Scale for Mexican Americans*) es una prueba psicológica que mide el grado de integración en la cultura anglosajona de los estadounidenses de origen mexicano en detrimento de la cultura mexicano-española. La distribución de los resultados de la prueba ARSMA en la población utilizada para desarrollar la prueba era aproximadamente normal, con una media de 3,0 y una desviación típica de 0,8. Un estudio posterior hizo pasar la prueba ARSMA a 42 estadounidenses hijos de padres de origen mexicano. La media de sus resultados fue $\bar{x} = 2{,}13$. Suponiendo que la desviación típica de esta población es también $\sigma = 0{,}8$, da un intervalo de confianza del 95% para la media de los resultados de la prueba ARSMA de los estadounidenses con padres de origen mexicano.

5.20. ¿Están satisfechos los directores de hotel con los sistemas informáticos que utilizan en sus hoteles? Se envió una encuesta a 560 directores de hoteles de entre 200 y 500 habitaciones en Chicago y Detroit,[3] de las cuales 135 fueron rellenadas. Dos preguntas de la encuesta hacían referencia a la facilidad de uso de los sistemas informáticos y al nivel de formación informática que habían recibido. Los directores respondieron utilizando una escala de 7 puntos, siendo 1 "no satisfecho", 4 "moderadamente satisfecho" y 7 "muy satisfecho".

(a) ¿Cuál crees que es la población de este estudio? Hay algunas deficiencias en la obtención de los datos. ¿Cuáles son? Estas deficiencias disminuyen el valor de la inferencia que estás a punto de hacer.

(b) En relación a la facilidad de uso de los sistemas informáticos, la media de los resultados fue $\bar{x} = 5{,}396$. Da un intervalo de confianza del 95% para la media poblacional (supón que la desviación típica poblacional es $\sigma = 1{,}75$).

(c) Por lo que respecta a la satisfacción con el nivel de formación informática recibido, la media de los resultados fue $\bar{x} = 4{,}398$. Tomando $\sigma = 1{,}75$, da un intervalo de confianza del 99% para la media poblacional.

(d) Las mediciones de la satisfacción no están, por supuesto, distribuidas normalmente, debido a que éstas sólo toman valores enteros entre 1 y 7. Sin embargo, la utilización de los intervalos de confianza basados en la distribución normal está justificada en este estudio. ¿Por qué?

5.21. Los consumidores de EE UU pueden adquirir medicamentos que no precisan receta en tiendas de alimentación, en grandes superficies o en farmacias. Aproximadamente un 45% de los consumidores hace estas compras en farmacias. ¿Qué hace que las farmacias vendan más, si frecuentemente tienen precios más altos?

Un estudio examinó la percepción de los consumidores sobre el desempeño de los tres tipos de establecimientos, utilizando un extenso cuestionario que preguntaba cosas tales como "tienda agradable y atractiva", "personal con conocimientos" y "ayuda en la elección entre varios tipos de medicamentos sin receta". El resultado del estudio se basaba en 27 preguntas de ese tipo. Los sujetos fueron 201 personas escogidas al azar de la guía telefónica de Indianapolis. He aquí las medias y las desviaciones típicas de las puntuaciones otorgadas por los miembros de la muestra.[4]

Tipo de tienda	$\bar{x}$	s
Tiendas de alimentación	18.67	24,95
Grandes superficies	32.38	33,37
Farmacias	48.60	35,62

No conocemos las desviaciones típicas de la población, pero una desviación típica muestral s extraída de una muestra tan grande normalmente se acerca a σ. En este ejercicio, utiliza s en lugar de la σ desconocida.

(a) ¿De qué población crees que los autores del estudio quieren extraer las conclusiones? ¿De qué población estás seguro que se pueden extraer conclusiones?

(b) Da intervalos de confianza del 95% para la media del desempeño de cada tipo de tienda.

(c) Basándote en los intervalos de confianza, ¿estás convencido de que los consumidores creen que las farmacias son mejores que los otros tipos de tiendas?

5.22. El cuestionario de formato ampliado del censo estadounidense de 1990 fue enviado a una muestra aleatoria del 17% de los hogares de este país. Una pregunta de este cuestionario hacía referencia a los ingresos totales del cabeza de familia en 1989 (el cabeza de familia es la persona a cuyo nombre está o se alquila la vivienda). Supón que los hogares que rellenaron el cuestionario ampliado son una muestra aleatoria simple de la población de todos los hogares de cada distrito. En Middletown, una ciudad de 40.000 habitantes, 2.621 cabezas de familia dijeron cuáles eran sus ingresos. La media de las respuestas fue $\bar{x} = 23.453$ dólares y la desviación típica fue $s = 8.721$. La desviación típica muestral para una muestra tan grande estará muy próxima a la desviación típica poblacional σ. Utiliza esto para dar un intervalo de confianza del 99% para la media de los ingresos de los cabezas de familia de Middletown en 1989.

5.23. ¿Qué tamaño de muestra de los hogares te permitiría estimar la media de los ingresos de los cabezas de familia de Middletown con un error de estimación de 1.000 dólares y con una confianza del 99%? (Repasa el ejercicio anterior).

5.24. Una encuesta del *New York Times* sobre temas de interés para la mujer entrevistó a 1.025 mujeres y a 472 hombres seleccionados aleatoriamente en EE UU, exceptuan-

[3] Datos proporcionados por John Rousselle y Huei-Ru Shieh, Department of Restaurant, Hotel, and Institutional Management, Purdue University.

[4] Datos proporcionados por Mugdha Gore y Joseph Thomas, Facultad de Farmacia, Purdue University.

do Alaska y Hawai. La encuesta daba un error de estimación de ± 3 puntos porcentuales con una confianza del 95% para los resultados de las mujeres. El error de estimación de los resultados de los hombres era de ± 4 puntos porcentuales. ¿Por qué el error de estimación de los hombres es mayor que el error de estimación de las mujeres?

5.25. Cuando el estadístico que estima un parámetro desconocido tiene una distribución normal, un intervalo de confianza para el parámetro tiene la forma

$$\text{estimación} \pm z^{*}\sigma_{\text{de la estimación}}$$

En un diseño de encuesta muestral complejo, la estimación de la media y de la desviación típica poblacional de esta estimación exige cálculos elaborados. Pero cuando se nos da la estimación y su desviación típica, podemos calcular un intervalo de confianza para μ sin conocer las fórmulas que condujeron a estos valores.

Un informe basado en la Encuesta Mensual de la Población Activa de EE UU estima que la mediana de los ingresos semanales de los trabajadores asalariados es de 664 dólares y que la desviación típica de esta estimación es de 3,50 dólares. La Encuesta Mensual de la Población Activa utiliza un elaborado diseño muestral en etapas múltiples para seleccionar una muestra de aproximadamente 60.000 hogares. La distribución de la mediana de los ingresos estimados es aproximadamente normal. Da un intervalo de confianza del 95% para la mediana de los ingresos semanales de todas las familias de trabajadores asalariados.

5.3 Pruebas de significación

Los intervalos de confianza son uno de los dos procedimientos de inferencia estadística más ampliamente utilizados. Utilízalos cuando tu objetivo sea estimar un parámetro poblacional. El segundo procedimiento de inferencia más ampliamente utilizado tiene otro objetivo: valorar la evidencia proporcionada por los datos a favor de alguna hipótesis sobre la población. Los razonamientos de las pruebas de significación, de forma análoga a los razonamientos de los intervalos de confianza, se basan en preguntar qué ocurriría si repitiéramos el muestreo o el experimento muchas veces. He aquí el primer ejemplo que exploraremos.

EJEMPLO 5.7

Los fabricantes de refrescos *light* utilizan edulcorantes artificiales con el objeto de evitar el azúcar. Los refrescos con este tipo de aditivos pierden poco a poco su sabor

dulce. En consecuencia, los industriales, antes de comercializar nuevos refrescos, determinan la pérdida de dulzor. Unos catadores experimentados valoran la dulzura de los refrescos, utilizando como referencia una serie de patrones, en una escala que va de 1 a 10. Posteriormente, se guardan los refrescos durante un mes a altas temperaturas para simular el efecto de un almacenamiento a temperatura ambiente durante 4 meses. Pasado este tiempo, los catadores vuelven a valorar la dulzura de los refrescos. Éste es un experimento por pares. Nuestros datos son las diferencias (la valoración inicial menos la valoración después del almacenamiento) entre las puntuaciones de los catadores. A mayor diferencia, mayor es la pérdida de dulzura.

He aquí las pérdidas de dulzura para un nuevo refresco, tal como las han determinado 10 catadores experimentados:

2,0 0,4 0,7 2,0 −0,4 2,2 −1,3 1,2 1,1 2,3

La mayor parte de los datos son positivos. Es decir, la mayoría de los catadores hallaron una pérdida de dulzura. De todas formas, las pérdidas son pequeñas e incluso dos catadores (las puntuaciones negativas) detectaron un incremento en la dulzura de los refrescos. *¿Constituyen estos datos una buena evidencia a favor de que los refrescos perdieron dulzura durante su almacenamiento?* ■

5.3.1 Razonamientos de las pruebas de significación

El promedio de la pérdida de dulzura de los refrescos viene dado por la media muestral,

$$\bar{x} = \frac{2{,}0 + 0{,}4 + \ldots + 2{,}3}{10} = 1{,}02\,.$$

Prueba de significación

No es una pérdida grande. Diez catadores distintos habrían dado, casi seguro, un resultado diferente. Puede ser que este resultado se deba sólo al azar. Una ***prueba de significación*** pregunta:

El resultado muestral $\bar{x} = 1{,}02$, ¿refleja una verdadera pérdida de dulzura?

O BIEN

El resultado $\bar{x} = 1{,}02$, ¿lo podemos obtener fácilmente sólo por azar?

Las pruebas de significación empiezan planteando estas dos alternativas. Nuestras conclusiones siempre se refieren a algún parámetro de la población. Por tanto, en primer lugar, tenemos que identificar este parámetro. En este caso se trata de la media poblacional μ. La media μ es el promedio de las pérdidas de dulzura que un número muy grande de catadores detectaría en los refrescos. Nuestros 10 catadores son una muestra de esta población.

Hipótesis nula

A continuación, hay que definir la ***hipótesis nula***. La hipótesis nula afirma que no hay ningún cambio o efecto en la población. Si la hipótesis nula es cierta, los resultados muestrales sólo indican una variación debida al azar. En nuestro ejemplo, la hipótesis nula afirma que la bebida no pierde dulzura (no hay cambio). Podemos escribir esto en términos de la pérdida media de dulzura μ de la población como

$$H_0 : \mu = 0$$

Escribimos H_0 para indicar la hipótesis nula.

Hipótesis alternativa

El efecto que sospechamos puede ser cierto, la alternativa al "efecto nulo" o al "cambio nulo", se describe como la ***hipótesis alternativa***. Creemos que la bebida pierde dulzura. En términos de la pérdida media de dulzura μ, la hipótesis alternativa es

$$H_a : \mu > 0$$

El razonamiento que hay detrás de las pruebas de significación es el siguiente:

- Supón que la hipótesis nula sea cierta, es decir, que en promedio no haya pérdida de dulzura.
- *¿Es el resultado muestral $\bar{x} = 1{,}02$ sorprendentemente grande bajo ese supuesto?* Si es así, existe evidencia en contra de H_0 y a favor de H_a.

Para responder a esta pregunta, utilizamos nuestro conocimiento sobre cómo variaría la media muestral $\bar{x}$ en un muestreo repetido, si H_0 fuera realmente cierta. De nuevo nos encontramos con la distribución de $\bar{x}$.

Por experiencia sabemos que la distribución individual de las puntuaciones de los catadores varía de acuerdo con una distribución normal. La media de esta distribución es el parámetro μ. Preguntamos qué sucedería si en realidad no hubiera ningún cambio en la media de las puntuaciones, es decir, si $\mu = 0$. Esto es justamente lo que plantea la hipótesis nula. También por experiencia sabemos que la desviación

típica de las puntuaciones de todos los catadores individuales es $\sigma = 1$ (no es realista suponer que conocemos la desviación típica de la población σ. En el próximo capítulo prescindiremos de este supuesto). La distribución de $\bar{x}$ de 10 catadores, por tanto, es normal con media $\mu = 0$ y desviación típica

$$\frac{\sigma}{\sqrt{n}} = \frac{1}{\sqrt{10}} = 0{,}316$$

Podemos juzgar si algún valor concreto de $\bar{x}$ es sorprendente, situándolo en esta distribución. La figura 5.8 muestra la distribución muestral con los valores observados de $\bar{x}$ para dos tipos de refrescos.

- Un refresco tenía $\bar{x} = 0{,}3$ para una muestra de 10 catadores. La figura 5.8 deja claro que un valor de $\bar{x}$ como éste puede ocurrir fácilmente sólo por azar cuando la media poblacional es $\mu = 0$. Que 10 catadores hallen que $\bar{x} = 0{,}3$, no constituye una evidencia a favor de que se haya producido una pérdida de dulzura.

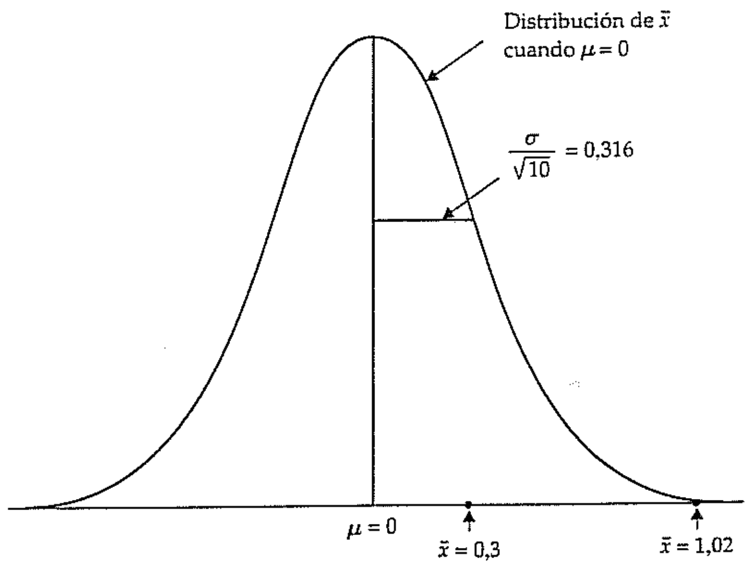


Figura 5.8. Si un refresco no pierde dulzura durante su almacenamiento, la puntuación media $\bar{x}$ de 10 catadores tendrá esta distribución muestral. El resultado de un refresco fue $\bar{x} = 0{,}3$. Esto podría ocurrir, fácilmente, por azar. Otro refresco tenía $\bar{x} = 1{,}02$. Este valor se encuentra muy alejado del valor $\mu = 0$ en la distribución normal y constituye, por tanto, una buena evidencia a favor de que dicho refresco ha perdido dulzura.

- La prueba de sabor de nuestro refresco dio $\bar{x} = 1.02$. Este valor de $\bar{x}$ queda muy alejado de $\mu = 0$ en la curva de la figura 5.8, tan alejado que no ocurriría casi nunca sólo por azar si el verdadero valor de μ fuera 0. Este valor observado constituye una buena evidencia a favor de que en realidad la verdadera μ es mayor que 0. Es decir, de que el refresco ha perdido dulzura. El fabricante debe reformular el refresco y probar otra vez.

Una prueba de significación consiste en preguntar cuál tendría que ser la probabilidad de un resultado observado si la hipótesis nula fuera realmente cierta. El paso final en nuestra prueba consiste en calcular un número que nos indique la probabilidad de que ocurra el valor observado $\bar{x}$ si H_0 es cierta. Cuanto menos probable sea este resultado, mayor es la evidencia en contra de H_0.

Valor P

Mira otra vez la figura 5.8. Si la hipótesis alternativa es cierta, hay una pérdida de dulzura y, por tanto, creemos que la pérdida media de dulzura $\bar{x}$ hallada por los catadores será positiva. Cuanto más alejada se encuentre $\bar{x}$ en la dirección positiva, más convencidos estaremos de que la media poblacional μ no es cero sino que toma un valor positivo. Medimos la fuerza de la evidencia en contra de H_0 por la probabilidad por debajo de la curva normal de la figura 5.8 situada a la derecha del valor observado $\bar{x}$. Esta probabilidad se llama ***valor P***. Es la probabilidad de que un resultado se encuentre al menos tan alejado de μ como el valor observado. Cuanto menor sea esta probabilidad, más sorprendente será nuestro resultado, y más fuerte será la evidencia en contra de la hipótesis nula.

- Para el nuevo refresco, nuestros 10 catadores dieron una $\bar{x} = 0.3$. La figura 5.9 muestra el valor P de este resultado. Es la probabilidad situada a la derecha de 0.3. Esta probabilidad es aproximadamente 0.17. Es decir, cuando la verdadera media poblacional es 0, un 17% de las muestras darían una puntuación media tan grande o mayor que 0.3 sólo por azar. Un resultado con esta probabilidad de ocurrir sólo por azar no constituye una buena evidencia en contra de la hipótesis nula.
- Nuestra bebida dio una mayor pérdida de dulzura, $\bar{x} = 1.02$. La probabilidad de un resultado como éste o mayor es sólo 0.0006. Esta probabilidad es el valor P. Diez catadores darían una puntuación media al menos tan grande como 1.02 sólo 6 veces de 10.000, si el verdadero cambio medio de dulzura fuese 0. Un resultado tan poco probable nos convence de que la verdadera media es realmente mayor que 0.

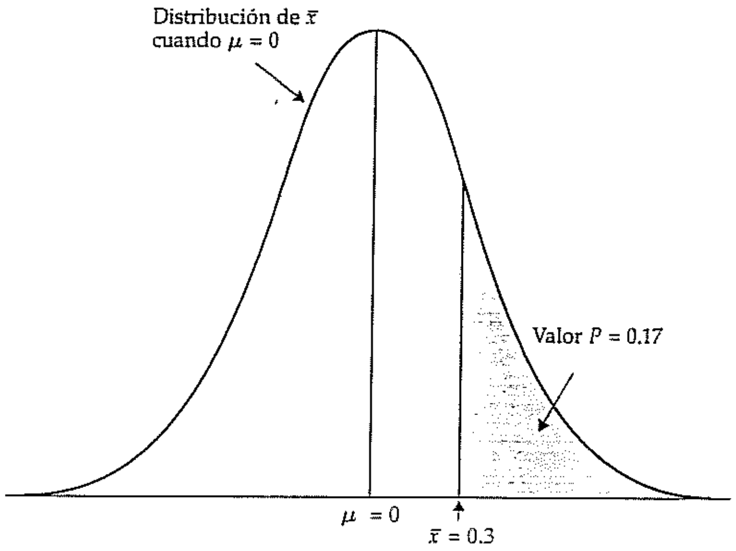


Figura 5.9. Valor P para el resultado $\bar{x} = 0.3$ en la cata de un refresco. El valor P es la probabilidad (cuando H_0 es verdadera) de que $\bar{x}$ tome un valor al menos tan grande como el valor realmente observado.

Significación estadística

Los valores P pequeños constituyen una evidencia en contra de H_0, ya que indican que el valor observado es poco probable que ocurra sólo por azar. Los valores P grandes no constituyen ninguna evidencia en contra de H_0. ¿Tiene que ser muy pequeño el valor P para que podamos considerar que existe suficiente evidencia en contra de H_0? No existe ninguna regla. En la práctica, se suele utilizar el nivel 0.05 (un resultado que, por azar, ocurriría no más de una vez de cada 20 intentos) como referencia. Un resultado con un valor P pequeño, digamos menor que 0.05, se considera ***estadísticamente significativo***. Es sólo una manera de decir que únicamente por azar difícilmente se produciría un resultado tan extremo.

5.3.2 Cómo se procede en una prueba de significación

He aquí, en forma resumida, los pasos que se siguen en una prueba de significación:

1. Describe el efecto que estás buscando en términos de un parámetro poblacional como la media μ (nunca plantees una hipótesis en términos de un estadístico muestral como $\bar{x}$).

2. La hipótesis nula es la afirmación de que este efecto *no* está presente en la población.

3. A partir de los datos, calcula un estadístico como $\bar{x}$ que estime el parámetro. ¿Está el valor de este estadístico muy alejado del valor del parámetro que establece la hipótesis nula? Si es así, los datos aportan evidencia a favor de que la hipótesis nula es falsa y de que, por tanto, el efecto que estás buscando realmente existe.

4. El valor P indica cuál sería la probabilidad de que ocurriera un resultado al menos tan extremo como el valor observado, suponiendo que la hipótesis nula fuera cierta. Los resultados con valores P pequeños ocurrirían raramente si la hipótesis nula fuera cierta. A estos resultados los denominamos estadísticamente significativos.

Esta descripción pasa por alto muchos detalles y muchos aspectos sutiles de las pruebas de significación. Pero es importante tener este esquema bien presente antes de seguir adelante. Las fórmulas de las pruebas de significación no muestran el razonamiento subyacente. De hecho, los programas estadísticos a menudo sólo dan el valor P. Mira otra vez la figura 5.8 y fíjate en los dos valores de $\bar{x}$ correspondientes a las pruebas de cata de dos refrescos. Te tiene que quedar claro que uno de los resultados no es sorprendente, si la verdadera media poblacional es 0, y que el otro sí lo es. Una prueba de significación simplemente dice esto de manera más precisa.

EJERCICIOS

5.26. La prueba SSHA (*Survey of Study Habits and Attitudes*) es una prueba psicológica que mide la actitud hacia la escuela y los hábitos de estudio de los alumnos. Los resultados van de 0 a 200. El resultado medio de los estudiantes universitarios de EE UU es aproximadamente 115 y la desviación típica aproximadamente 30. Una profesora intuye que los estudiantes de más edad tienen una mejor actitud hacia la escuela. La profesora hace pasar la prueba SSHA a 25 estudiantes que tienen como mínimo 30 años. Supón que los resultados de la población de estudiantes mayores de 30 años se distribuye normalmente con desviación típica $\sigma = 30$. La profesora quiere contrastar las hipótesis

$$H_0 : \mu = 115$$
$$H_a : \mu > 115$$

(a) ¿Cuál es la distribución de la media de los resultados $\bar{x}$ de una muestra de 25 estudiantes mayores de 30 años si la hipótesis nula es cierta? Dibuja la curva de densidad de esta distribución (sugerencia: primero dibuja una curva normal, luego, a partir de la información recibida, señala en el eje de las abcisas la localización de μ y de σ en la curva normal).

(b) Supón que los datos de la muestra dan $\bar{x} = 118{,}6$. Señala este punto en el eje de las abcisas de tu gráfico. En realidad, el resultado fue $\bar{x} = 125{,}7$. Señala este punto en tu dibujo. Utilizando tu dibujo, explica con un lenguaje sencillo por qué uno de los resultados constituye una buena evidencia a favor de que la media de los resultados de todos los estudiantes mayores de 30 años es mayor que 115 y por qué el otro resultado no lo es.

(c) Sombrea el área por debajo de la curva que corresponde al valor P del resultado muestral $\bar{x} = 118{,}6$.

5.27. La Oficina del Censo de EE UU informa que los hogares de este país dedicaron un promedio del 31% de todos sus gastos a la vivienda. Una asociación de constructores de la ciudad de Cleveland cree que este promedio es menor en su zona. Estos constructores entrevistan una muestra de 40 hogares del área metropolitana de Cleveland para saber qué porcentaje de sus gastos se dedica a la vivienda. Sea μ la media de los porcentajes del gasto dedicado a la vivienda en los hogares de Cleveland. Queremos contrastar las hipótesis

$$H_0 : \mu = 31\%$$
$$H_a : \mu < 31\%$$

La desviación típica de la población es $\sigma = 9{,}6\%$.

(a) ¿Cuál es la distribución de la media de los porcentajes $\bar{x}$ del gasto que las muestras dedican a la vivienda si la hipótesis nula es cierta? Dibuja la curva de densidad de la distribución muestral. (Sugerencia: primero dibuja una curva normal, luego señala en el eje de las abcisas lo que sabes sobre la posición de μ y de σ en una curva normal).

(b) Supón que el estudio halla que $\bar{x} = 30{,}2\%$ para los 40 hogares de la muestra. Señala este punto en el eje de las abcisas de tu dibujo. Luego supón que el resultado del estudio es $\bar{x} = 27{,}6\%$. Señala este punto en tu dibujo. Refiriéndote a tu dibujo, explica con un lenguaje sencillo por qué uno de los resultados constituye una buena evidencia a favor de que el promedio del gasto en vivienda en los hogares de Cleveland es menor que el 31%, y por qué el otro resultado no lo es.

(c) Sombrea el área por debajo de la curva que da el valor P para el resultado de $\bar{x} = 30{,}2\%$ (fíjate en que estamos buscando evidencia a favor de que el gasto es *menor* que el que supone la hipótesis nula).

5.3.3 Más detalles: planteamiento de las hipótesis

Ahora vamos a ver con más detalle algunos aspectos de las pruebas de significación. Lo primero en una prueba de significación consiste en hacer una afirmación *contra* la cual intentaremos encontrar evidencia.

> **HIPÓTESIS NULA H_0**
>
> La afirmación que se contrasta en una prueba de significación se llama **hipótesis nula**. Las pruebas de significación se diseñan con el objetivo de valorar la fuerza de la evidencia en contra de la hipótesis nula. En general, la hipótesis nula es una afirmación de "ausencia de efecto" o de "no diferencia".

La hipótesis alternativa H_a es una afirmación sobre la población *a favor* de la cual intentamos encontrar evidencia. En el ejemplo 5.7 estuvimos buscando evidencia a favor de una pérdida de dulzura. La hipótesis nula establece que "no hay pérdida" de dulzura en promedio en una población de catadores grande. La hipótesis alternativa establece que "sí hay pérdida". Por tanto, las hipótesis son

$$H_0 : \mu = 0$$
$$H_a : \mu > 0$$

Pruebas de significación de una cola

Esta hipótesis alternativa es de ***una cola***, ya que sólo estamos interesados en desviaciones de la hipótesis nula en una dirección.

EJEMPLO 5.8

El grado de satisfacción en el trabajo de los operarios de líneas de montaje, ¿es distinto en función de si el ritmo de trabajo lo marcan ellos mismos o de si éste lo marca una máquina? Un estudio escogió al azar 28 sujetos de un grupo de operarias de una línea de montaje de componentes electrónicos. Este grupo se dividió, también al azar, en dos mitades. Una mitad fue asignada a una línea de montaje con el ritmo de trabajo marcado por una máquina y la otra mitad a una línea de montaje de características similares a la anterior, pero en la cual las operarias marcaban su propio ritmo de trabajo. Al cabo de dos semanas todas las operarias pasaron una prueba para determinar su grado de satisfacción en el trabajo. Después de pasar la prueba, los dos subgrupos se intercambiaron. Dos semanas más tarde, las operarias volvieron a pasar la prueba. Esta experiencia constituye otro ejemplo de diseño por pares. La variable respuesta es la diferencia entre las puntuaciones después de estar trabajando en la línea al ritmo de trabajo marcado por las operarias y después de estar trabajando en la línea de montaje al ritmo de trabajo marcado por una máquina.[5]

Pruebas de significación de dos colas

El parámetro de interés es la media μ de las diferencias entre los resultados de las dos pruebas en la población de todas las operarias en líneas de montaje. La hipótesis nula dice que no hay diferencias entre las dos condiciones de trabajo, es decir,

$$H_0 : \mu = 0$$

Los autores del estudio querían saber si las dos condiciones de trabajo proporcionaban satisfacciones distintas a las operarias. Los autores no especificaron la dirección de la diferencia. Por tanto, la hipótesis alternativa es de ***dos colas***,

$$H_a : \mu \neq 0 \quad \blacksquare$$

Las hipótesis siempre se refieren a alguna población, no a resultados particulares. Por este motivo, plantea siempre H_0 y H_a en términos de los parámetros poblacionales. Debido a que H_a expresa el efecto del cual esperamos encontrar evidencia *a favor*, a menudo es más fácil empezar planteando H_a y luego plantear H_0 como la hipótesis de que la evidencia a favor no está presente.

No siempre es fácil decidir si H_a tiene que ser de una o de dos colas. En el ejemplo 5.8 la alternativa $H_a : \mu \neq 0$ es de dos colas. Afirma, simplemente, que hay diferencias en el grado de satisfacción sin especificar la dirección de la diferencia. La alternativa $H_a : \mu > 0$ en el ejemplo de la prueba de cata es de una cola. Debido a que los refrescos sólo pueden perder dulzor durante el almacenamiento, sólo estamos interesados en detectar variaciones positivas de μ. La hipótesis alternativa debe expresar nuestras sospechas o nuestra esperanza. Es hacer trampa mirar primero los datos y a continuación plantear la H_a que mejor se ajuste a ellos. Así, por ejemplo, el hecho de que las operarias en el estudio del ejemplo 5.8 estuvieran más satisfechas cuando ellas mismas marcaban el ritmo de trabajo no tiene que influir en la elección

[5] G. Salvendy, G. P. McCabe, S. G. Sanders, J. L. Knight y E. J. McCormick, 1982, "Impact of personality and intelligence on job satisfaction of assembly line and bench work —an industrial study", *Applied Ergonomics*, 13, págs. 293-299.

de H_a. Si no has tomado por adelantado una decisión firme sobre la dirección del efecto, utiliza una alternativa de dos colas.

La elección de las hipótesis del ejemplo 5.7 como

$$H_0 : \mu = 0$$
$$H_a : \mu > 0$$

merece un comentario final. El fabricante de refrescos no está interesado en la posibilidad de que se produzca una ganancia de dulzura, que sería indicada con un valor negativo de la pérdida media μ. De todas formas, podríamos dejar abierta la posibilidad de que μ fuera menor que 0 incluyendo este caso en la hipótesis nula. Entonces escribiríamos:

$$H_0 : \mu \leq 0$$
$$H_a : \mu > 0$$

Este planteamiento es satisfactorio desde un punto de vista lógico, ya que las hipótesis cubren todos los valores de μ. Sin embargo, sólo el valor del parámetro bajo H_0 que se encuentra más cerca de H_a influye sobre la forma de la prueba en todos los casos habituales en los que se contrastan hipótesis. Por tanto, escribiremos H_0 de la forma más sencilla posible, es decir, aquella en la que el parámetro tiene un valor concreto, en este caso $H_0 : \mu = 0$.

EJERCICIOS

Todas las situaciones que se plantean a continuación se pueden resolver mediante una prueba de significación para una media poblacional μ. En cada caso, plantea las hipótesis nula H_0 y alternativa H_a.

5.28. El diámetro de un eje de un pequeño motor se supone que es de 5 mm. Si el eje es demasiado pequeño o demasiado grande, el motor no funcionará adecuadamente. El fabricante mide el diámetro de los ejes de una muestra de motores para determinar si el diámetro medio se ha desviado del objetivo.

5.29. Los datos de la Oficina Censal de EE UU muestran que la media de los ingresos de los hogares en el área de influencia de un gran centro comercial es de 42.500 dólares por año. Una empresa de investigación de mercados entrevista a los clientes del centro comercial. Los investigadores sospechan que la media de los ingresos de estos clientes es mayor que la media de los ingresos de la población general.

5.30. La nota media de contabilidad de un gran grupo de estudiantes es 5. El catedrático cree que uno de sus ayudantes no es muy bueno y sospecha que los alumnos de ese profesor ayudante tienen una nota media menor que la general del grupo. Los estudiantes de ese profesor ayudante pueden considerarse una muestra de la población de todos los estudiantes del curso, así que el catedrático compara la media de esos estudiantes con la media general del grupo.

5.31. El año pasado el servicio técnico de tu empresa dedicó un promedio de 2.6 horas a resolver por teléfono los problemas de los clientes de la empresa con contratos de servicio. Los datos de este año, ¿muestran un promedio de tiempo dedicado a resolver problemas por teléfono distinto?

5.3.4 Más detalles: valores P y significación estadística

Estadístico de contraste

Una prueba de significación utiliza los datos en forma de *estadístico de contraste*. El estadístico de contraste se basa, normalmente, en un estadístico que estima el parámetro que aparece en las hipótesis. En nuestros ejemplos, el parámetro es μ y el estadístico de contraste es la media muestral $\bar{x}$.

Una prueba de significación valora la evidencia en contra de la hipótesis nula en términos de probabilidad. Si el estadístico de contraste se sitúa lejos del valor propuesto en la hipótesis nula y en la dirección expresada por la hipótesis alternativa, esto constituye una buena evidencia en contra de H_0 y a favor de H_a. Para describir la fuerza de la evidencia, halla la probabilidad de obtener *un resultado al menos tan extremo como el resultado observado*. "Extremo" significa "lejos del valor que esperaríamos si H_0 fuera cierta". La dirección o direcciones que se tienen en cuenta en "lejos del valor que esperaríamos" están determinadas por la hipótesis alternativa H_a.

VALOR P

La probabilidad, suponiendo que H_0 es cierta, de que el estadístico de contraste tome un valor tan o más extremo que el valor realmente observado se llama valor P de la prueba. Cuanto menor sea el valor P, más fuerte es la evidencia que proporcionan los datos en contra de H_0.

La mayoría de los programas estadísticos que llevan a cabo pruebas de significación calculan y dan los valores P. En algunos casos podemos calcular el valor P si conocemos la distribución muestral del estadístico de contraste.

EJEMPLO 5.9

En el ejemplo 5.7 las observaciones son una muestra aleatoria simple de tamaño $n = 10$ de una población normal con $\sigma = 1$. La pérdida de dulzura media observada en un refresco fue $\bar{x} = 0.3$. El valor P para contrastar

$$H_0 : \mu = 0$$
$$H_a : \mu > 0$$

es, por tanto,

$$P(\bar{x} \geq 0.3)$$

calculado suponiendo que H_0 es cierta. Cuando H_0 es cierta, $\bar{x}$ tiene una distribución normal de media $\mu = 0$ y desviación típica igual a

$$\frac{\sigma}{\sqrt{n}} = \frac{1}{\sqrt{10}} = 0.316$$

Halla el valor P a partir del cálculo de probabilidades normales. Empieza dibujando la distribución de $\bar{x}$ y sombreando el área correspondiente al valor P. La figura 5.10 es el gráfico de este ejemplo. Luego estandariza $\bar{x}$ para tener la distribución normal estandarizada Z y utiliza la tabla A,

$$P(\bar{x} \geq 0.3) = P\left(\frac{\bar{x} - 0}{0{,}316} \geq \frac{0{,}3 - 0}{0{,}316}\right) =$$
$$= P(Z \geq 0{,}95) = 1 - 0{,}8289 = 0{,}1711$$

Este valor P es el que aparece en la figura 5.9. ■

Nivel de significación

Algunas veces damos un último paso para valorar la evidencia en contra de H_0. Comparamos el valor P con un valor previamente determinado que consideramos decisivo. Esto equivale a decidir de antemano cuál consideramos que tiene que ser la evidencia en contra de H_0. El valor decisivo de P se llama ***nivel de significación***. Lo simbolizamos como α, la letra griega alfa. Si escogemos $\alpha = 0{,}05$, exigimos que los datos proporcionen una evidencia en contra de H_0 tan fuerte que el valor de $\bar{x}$ no ocurriría por azar más del 5% de las veces (1 vez de cada 20) cuando H_0 fuera cierta. Si escogemos $\alpha = 0.01$, exigimos una evidencia aún más fuerte en contra de H_0, una evidencia tan fuerte que el valor de $\bar{x}$ sólo ocurriría por azar en un 1% de las ocasiones (1 vez de cada 100) si H_0 fuera cierta.

> **SIGNIFICACIÓN ESTADÍSTICA**
>
> Si el valor P es más pequeño o igual que α, decimos que los datos son estadísticamente significativos a un nivel α.

"Significativo" en estadística no quiere decir "importante". Simplemente quiere decir que "es muy poco probable que ocurra sólo por azar". El nivel de significación α hace que "poco probable" sea más preciso. Significativo a un nivel 0.01 se expresa, a menudo, de la siguiente manera: "los resultados eran significativos ($P < 0.01$)". Aquí P simboliza el valor P. El valor P aporta más información que la afirmación sobre la significación, ya que entonces podemos valorar la significación a cualquier nivel que escojamos. Por ejemplo, un resultado con $P = 0.03$ es significativo a un nivel de significación $\alpha = 0{,}05$, pero no lo es a un nivel de significación $\alpha = 0{,}01$.

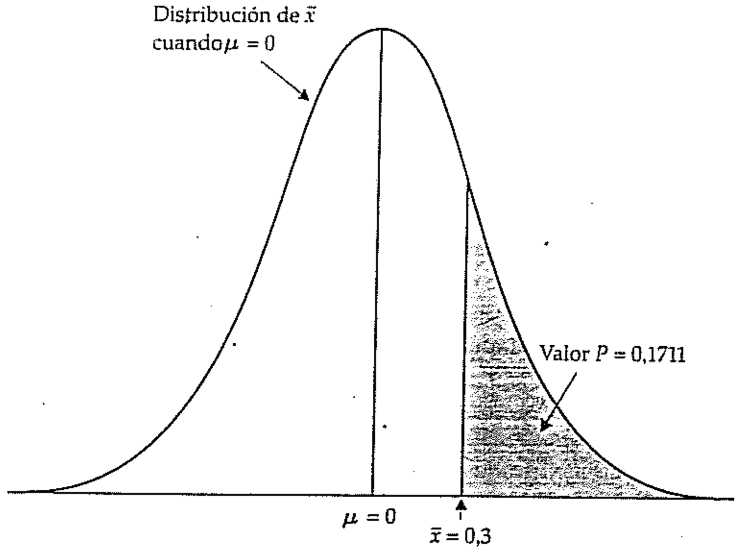


Figura 5.10. Valor P en la prueba de una cola del ejemplo 5.9.

EJERCICIOS

5.32. Vuelve al ejercicio 5.26. A partir de tu dibujo, calcula los valores P de $\bar{x} = 118.6$ y de $\bar{x} = 125.7$. Los dos valores P expresan en cifras la comparación que hiciste informalmente en el ejercicio 5.26.

5.33. Vuelve al ejercicio 5.27. A partir de tu dibujo, calcula los valores P de $\bar{x} = 30.2\%$ y de $\bar{x} = 27.6\%$. Los dos valores P expresan en cifras la comparación que hiciste informalmente en el ejercicio 5.27.

5.34. Las ventas semanales de café molido en un supermercado tienen una distribución normal con una media $\mu = 354$ unidades por semana y una desviación típica de $\sigma = 33$ unidades. La tienda reduce el precio del café en un 5%. Las ventas en las tres semanas siguientes son de 405, 378 y 411 unidades. Estos resultados, ¿constituyen una buena evidencia a favor de que el promedio de ventas es ahora mayor? Las hipótesis son:

$$H_0 : \mu = 354$$
$$H_a : \mu > 354$$

Supón que la desviación típica de la población de las ventas semanales se mantiene en $\sigma = 33$.

(a) Halla el valor del estadístico de contraste $\bar{x}$.

(b) Dibuja la curva normal de la distribución de $\bar{x}$ cuando H_0 es cierta. Sombrea el área que representa el valor P del resultado observado.

(c) Calcula el valor P.

(d) El resultado, ¿es estadísticamente significativo al nivel $\alpha = 0{,}05$? ¿Es significativo al nivel $\alpha = 0.01$? ¿Crees que existe suficiente evidencia a favor de que la media de ventas es mayor?

5.35. Un estudio sobre el salario de los altos ejecutivos examinó, en un año reciente, el aumento de los ingresos de los ejecutivos de 104 empresas teniendo en cuenta la inflación. El incremento medio de los ingresos fue $\bar{x} = 6{,}9\%$ y la desviación típica de los incrementos fue $s = 55\%$. Estos resultados, ¿constituyen una buena evidencia a favor de que el ingreso medio μ de todos los altos ejecutivos aumentó ese año? Las hipótesis son:

$$H_0 : \mu = 0 \text{ (sin incremento)}$$
$$H_a : \mu > 0 \text{ (con incremento)}$$

Debido a que el tamaño de la muestra es grande, la s muestral está próxima a la σ poblacional, por tanto, considera que $\sigma = 55\%$.

(a) Dibuja la curva normal de la distribución de $\bar{x}$ cuando H_0 es verdadera. Sombrea el área que representa el valor P del resultado observado $\bar{x} = 6{,}9\%$.

(b) Calcula el valor P.

(c) El resultado, ¿es significativo al nivel $\alpha = 0{,}05$? ¿Crees que el estudio proporciona suficiente evidencia a favor de que subieron los ingresos de todos los altos ejecutivos?

5.36. Un sociólogo dice que "en nuestra muestra, el etnocentrismo fue significativamente mayor ($P < 0{,}05$) entre los asistentes a misa que entre los no asistentes". Explica lo que esto significa con un lenguaje comprensible para alguien que no sepa estadística. No utilices la palabra "significación" en tu respuesta.

5.37. La oficina de ayuda financiera de una universidad pregunta a una muestra de estudiantes sobre sus empleos y ganancias. El informe dice que "en relación a las ganancias de los estudiantes durante el curso académico, se halló una diferencia significativa ($P = 0{,}038$) entre sexos; en promedio, los hombres ganaron más que las mujeres. No se halló ninguna diferencia ($P = 0{,}476$) entre las ganancias de los estudiantes blancos y los estudiantes negros". Explica, con un lenguaje comprensible para alguien que no sepa estadística, estas dos conclusiones en relación a la influencia del sexo y de la raza sobre la media de las ganancias.[6]

5.3.5 Pruebas de significación para una media poblacional

Aunque la argumentación de las pruebas de significación no es sencilla, llevar a cabo una prueba sí lo es. Tienes que seguir tres pasos:

1. Plantea las hipótesis.
2. Calcula el estadístico de contraste.
3. Halla el valor P.

Una vez hayas formulado tus hipótesis e identificado el estadístico de contraste adecuado, puedes llevar a cabo los pasos 2 y 3 a mano o con tu ordenador. Vamos a desarrollar, ahora, el procedimiento que se debe seguir en una prueba de significación; el mismo procedimiento que hemos utilizado en nuestros ejemplos.

Tenemos una muestra aleatoria simple de tamaño n de una población normal de media desconocida μ. Queremos contrastar la hipótesis de que μ tiene un determi-

[6] De un estudio de M. R. Schlatter, *et al.*, División de Ayuda Financiera, Purdue University.

nado valor. Llama a este valor μ_0. La hipótesis nula es

$$H_0 : \mu = \mu_0$$

La prueba se basa en la media muestral $\bar{x}$. Debido a que el cálculo de probabilidades de distribuciones normales exige variables estandarizadas, utilizaremos como estadístico de contraste la media muestral estandarizada

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

Estadístico z

Este ***estadístico*** z tiene una distribución normal estandarizada cuando H_0 es cierta. Si la hipótesis alternativa es de una cola, la cola de la derecha,

$$H_a : \mu > \mu_0$$

entonces el valor P es la probabilidad de que una variable normal estandarizada Z tome un valor al menos tan grande como el valor observado z. Esto es,

$$P = P(Z \geq z)$$

El ejemplo 5.9 calcula este valor P en la prueba del refresco. Allí, $\mu_0 = 0$, la media muestral estandarizada era $z = 0{,}95$ y el valor P era $P(Z \geq 0{,}95) = 0{,}1711$. Un razonamiento similar se aplica cuando la hipótesis alternativa establece que la verdadera μ es menor que el valor de la hipótesis nula μ_0 (en una prueba de una cola).

Cuando H_a establece que μ es sencillamente distinta de μ_0 (prueba de dos colas), para hallar evidencia en contra de la hipótesis nula se tienen en cuenta los valores de z que quedan lejos de 0 en cualquier dirección. El valor P es la probabilidad de que la variable normal estandarizada Z se encuentre, al menos, tan lejos de cero en *cualquier dirección* como el valor z observado.

EJEMPLO 5.10

Supón que el valor del estadístico de contraste z para una prueba de dos colas es $z = 1{,}7$. El valor P de dos colas es la probabilidad de que $Z \leq -1{,}7$ o de que $Z \geq 1{,}7$. La figura 5.11 muestra esta probabilidad en forma de áreas por debajo de la curva normal estandarizada. Debido a que la distribución normal estandarizada es simétrica, podemos calcular esta probabilidad hallando la $P(Z \geq 1{,}7)$ y *doblando* su valor.

$$P(Z \leq -1{,}7 \text{ o } Z \geq 1{,}7) = 2P(Z \geq 1{,}7) = 2(1 - 0{,}9554) = 0{,}0892$$

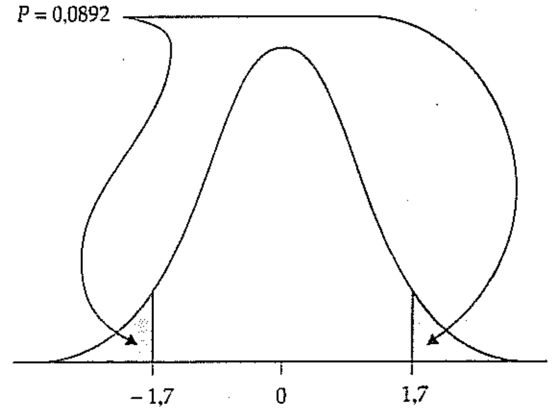


Figura 5.11. El valor P para la prueba de dos colas del ejemplo 5.10 es la suma de las áreas situadas a la izquierda de − 1.7 y a la derecha de 1.7.

Haríamos exactamente los mismos cálculos si el valor observado fuera $z = -1.7$. Lo que importa es el valor absoluto $|z|$, no si z es positivo o negativo. ■

EJEMPLO 5.11

Las autoridades sanitarias establecen que la presión sistólica media de la sangre de los hombres entre 35 y 44 años es 128 con una desviación típica igual a 15. El director médico de una gran empresa revisa sus archivos y halla que la presión sistólica media de una muestra de 72 ejecutivos de la empresa de este grupo de edad es $\bar{x} = 126.07$. ¿Se puede afirmar que la presión sistólica media de los ejecutivos de la empresa es distinta que la media poblacional? Como siempre en este capítulo, hacemos el supuesto, poco realista, de que conocemos la desviación típica poblacional. Supón que los ejecutivos tienen la misma $\sigma = 15$ que la población de todos los hombres adultos de mediana edad.

Paso 1: Hipótesis. La hipótesis nula establece que "no hay diferencias" con la media nacional $\mu_0 = 128$. La hipótesis alternativa es de dos colas, ya que el director médico no piensa en una dirección particular antes de examinar los datos. Por tanto, las hipótesis sobre la media desconocida μ de la población de ejecutivos son

$$H_0 : \mu = 128$$
$$H_a : \mu \neq 128$$

PRUEBA z PARA UNA MEDIA POBLACIONAL

Para contrastar la hipótesis $H_0: \mu = \mu_0$ a partir de una muestra aleatoria simple de tamaño n de una población con una media desconocida μ y una desviación típica σ conocida, calcula el **estadístico de contraste** z

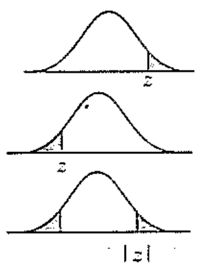
$$z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}$$

En términos de una variable Z que tiene una distribución normal estandarizada, los valores P para contrastar H_0 en contra de las siguientes alternativas

$H_a: \mu > \mu_0$, es $P = P(Z \geq z)$

$H_a: \mu < \mu_0$, es $P = P(Z \leq z)$

$H_a: \mu = \mu_0$, es $P = 2P(Z \geq |z|)$



Estos valores P son exactos si la distribución poblacional es normal y aproximadamente correctos para valores de n grandes, en los restantes casos.

Paso 2: Estadístico de contraste. El estadístico z de contraste es

$$z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} = \frac{126{,}07 - 128}{15/\sqrt{72}} = -1{,}09$$

Paso 3: Valor P. Sigue siendo conveniente que hagas un dibujo para ayudarte a hallar el valor P, aunque basta con que señales el valor z observado en una representación aproximada de la distribución normal estandarizada. La figura 5.12 muestra que el valor P es la probabilidad de que la variable normal estandarizada Z tome un valor que esté a una distancia de 0 de al menos 1,09. En la tabla A, hallamos que esta probabilidad es

$$P = 2P(Z \geq 1{,}09) = 2(1 - 0{,}8621) = 0{,}2758$$

Conclusión: Más de un 27% de las veces, una muestra aleatoria simple de tamaño 72 de la población de todos los hombres tendrá una presión sistólica de la sangre que esté situada al menos tan lejos de 128 como la media de la muestra de ejecutivos.

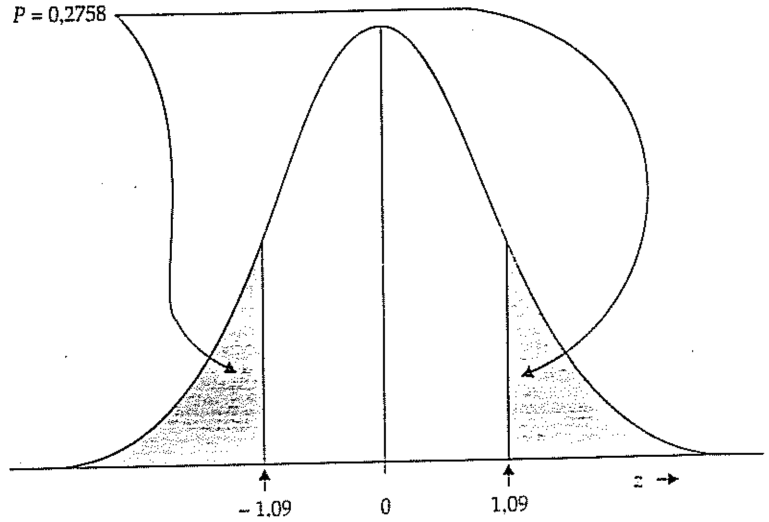


Figura 5.12. El valor P de la prueba de significación de dos colas del ejemplo 5.11.

La $\bar{x} = 126{,}07$ observada no constituye, por tanto, una buena evidencia a favor de que los ejecutivos difieren en este aspecto de los demás hombres. ■

El estadístico z supone que la muestra de 72 ejecutivos es una muestra aleatoria simple de la población de todos los ejecutivos varones de mediana edad de la empresa. Debemos comprobar este supuesto preguntando cómo se obtuvieron los datos. La muestra no sería de gran valor si sólo representara a los ejecutivos que, por ejemplo, en los últimos meses han tenido problemas de salud. Lo ideal sería que fuera una muestra aleatoria simple de todos los ejecutivos de la empresa. Si todos los ejecutivos tienen que pasar obligatoriamente cada año una revisión médica, no es difícil tomar una muestra aleatoria simple que represente al colectivo.

Los datos del ejemplo 5.11 *no* establecen que la media de la presión sistólica μ de los ejecutivos de la empresa, entre 35 y 44 años, sea igual a 128. Buscábamos evidencia a favor de que μ fuera distinta de 128 y no la encontramos. Esto es todo lo que podemos decir. Sin duda, la presión sistólica media de toda la población de ejecutivos no es exactamente igual a 128. Una muestra suficientemente grande nos proporcionaría información sobre la diferencia existente entre la media de la población general y la media de la población de ejecutivos, incluso si ésta fuera muy pequeña. Las pruebas de significación valoran la evidencia en *contra* de H_0. Si la evidencia en contra de H_0 es fuerte, podemos estar seguros de que hay que rechazar H_0 en favor de H_a. No encontrar evidencia en contra de H_0 sólo significa que los datos son compatibles con H_0, no que tengamos una clara evidencia a favor de que H_0 sea cierta.

EJEMPLO 5.12

En un debate sobre el nivel de formación de la mano de obra de EE UU, alguien dice, "en promedio, los jóvenes de hoy no son capaces ni de calcular el saldo de su cuenta corriente". El NAEP afirma que una persona que obtenga una puntuación de al menos 275 en la prueba de aritmética (repasa el ejemplo 5.2) está capacitada para determinar el saldo de una cuenta corriente. La muestra aleatoria de 840 hombres jóvenes de la encuesta NAEP tiene una puntuación media $\bar{x} = 272$, algo menor que la puntuación necesaria para calcular el saldo de una cuenta corriente. Estos resultados muestrales, ¿constituyen una buena evidencia a favor de que la media de *todos* los hombres jóvenes es menor que 275? Al igual que en el ejemplo 5.2, supón que $\sigma = 60$.

Paso 1: Hipótesis. Las hipótesis son

$$H_0 : \mu = 275$$
$$H_a : \mu < 275$$

Paso 2: Estadístico de contraste. El estadístico z es

$$z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} = \frac{272 - 275}{60/\sqrt{8400}} = -1{,}45$$

Paso 3: Valor P. Debido a que H_a es de una cola, la cola de la izquierda, son los valores pequeños de z los que se tienen en cuenta en contra de H_0. La figura 5.13 ilustra el valor P. Utilizando la tabla A, encontramos que

$$P = P(Z \leq -1{,}45) = 0{,}0735$$

Conclusión: Una puntuación tan pequeña como 272 ocurriría por azar, aproximadamente, en 7 de cada 100 muestras si la media poblacional fuera 275. Tenemos una cierta evidencia a favor de que la media de los resultados de los hombres jóvenes en la encuesta NAEP es menor que 275. ■

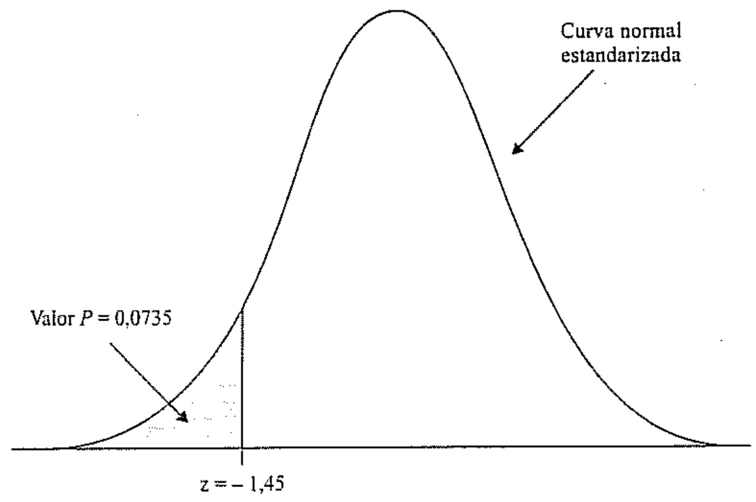


Figura 5.13. El valor P de la prueba de significación de una cola del ejemplo 5.12.

EJERCICIOS

5.38. He aquí las mediciones (en milímetros) de la dimensión crítica de una muestra de cigüeñales de motores de automóvil

224,120	224.001	224.017	223.982
223,960	224.089	223.987	223.976
224,098	224.057	223,913	223.999
223,989	223.902	223,961	223,980

Se sabe que el proceso de fabricación varía normalmente con una desviación típica $\sigma = 0{,}060$ mm. La media del proceso se supone que es de 224 mm. Estos datos, ¿proporcionan suficiente evidencia a favor de que la media del proceso no es igual al valor objetivo 224 mm?

(a) Plantea las H_0 y H_a que contrastarás.

(b) Calcula el estadístico z de contraste.

(c) Da el valor P de la prueba. ¿Estás convencido de que la media del proceso no es de 224 mm?

5.39. Se supone que las botellas de una famosa bebida refrescante contienen 300 mililitros (ml). Existe una cierta variación entre las botellas porque las máquinas embotelladoras no son absolutamente precisas. La distribución de los contenidos de las botellas es normal con una desviación típica $\sigma = 3$ ml. Un inspector que sospecha que la embotelladora llena menos de lo que debiera, mide el contenido de seis botellas. Los resultados son

299,4	297,7	301,0
298,9	300,2	297,0

Estos datos, ¿proporcionan suficiente evidencia a favor de que el contenido medio de las botellas de refresco es menor de 300 ml?

(a) Plantea las hipótesis que contrastarás.

(b) Calcula el estadístico de contraste.

(c) Halla el valor P y expresa tus conclusiones.

5.3.6 Pruebas con un nivel de significación predeterminado

Algunas veces exigimos un determinado grado de evidencia en contra de H_0 para rechazar la hipótesis nula. Un nivel de significación α establece la evidencia que exigimos. En términos de valores P, el resultado de una prueba de significación de nivel α es significativo si $P \leq \alpha$. Valorar la significación de una prueba es fácil cuando tienes el valor P. Cuando no utilizas un programa estadístico, el valor P puede ser difícil de calcular. Afortunadamente, puedes decidir si un resultado es estadísticamente significativo sin calcular P. El siguiente ejemplo ilustra cómo valorar la significación para un valor del nivel de significación α predeterminado utilizando una tabla de valores críticos, la misma tabla utilizada para obtener los intervalos de confianza.

EJEMPLO 5.13

En el ejemplo 5.12 examinábamos si la media de los resultados de los hombres jóvenes en la prueba de aritmética de la encuesta NAEP era menor que 275. Las hipótesis son

$$H_0 : \mu = 275$$
$$H_a : \mu < 275$$

El estadístico z toma el valor $z = -1{,}45$. La evidencia en contra de H_0, ¿es estadísticamente significativa a un nivel del 5%?

Para determinar la significación sólo tenemos que comparar el valor observado $z = -1{,}45$ con el valor crítico del 5%, $z^* = 1{,}645$ de la tabla C. Debido a que $z = -1{,}45$ *no* se encuentra más lejos de 0 que $-1{,}645$, la prueba *no* es significativa a un nivel $\alpha = 0{,}05$.

Vamos a ver por qué. El valor P es el área situada a la izquierda de $-1{,}45$ por debajo de la curva normal estandarizada que se muestra en la figura 5.13. El resultado $z = -1{,}45$ es significativo a un nivel del 5%, exactamente cuando esta área no es mayor que 0,05. El área situada a la izquierda del valor crítico $-1{,}645$ es exactamente 0,05. Por tanto, $-1{,}645$ separa los valores z que son significativos de los que no lo son. La figura 5.14 ilustra este procedimiento. ■

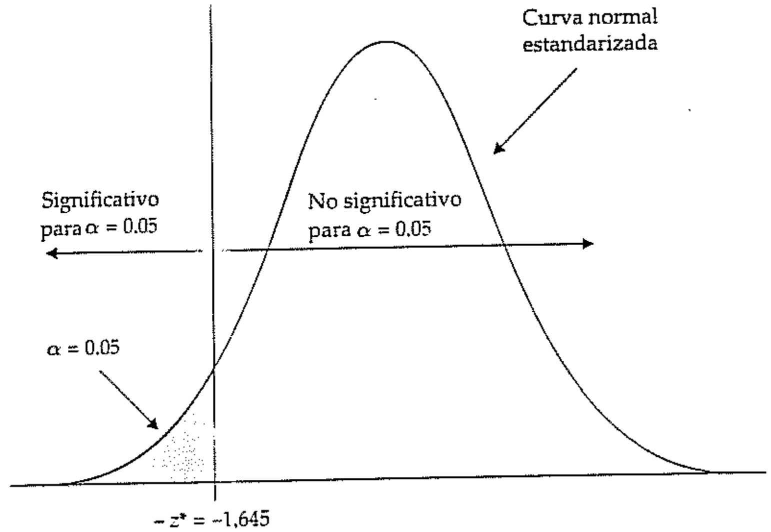


Figura 5.14. La decisión de si el estadístico z es significativo a un nivel $\alpha = 0{,}05$ en la prueba de una cola del ejemplo 5.13.

EJEMPLO 5.14

Al laboratorio de análisis del ejemplo 5.4 se le pide que evalúe si la concentración en materia activa de un comprimido es del 0.86%. El laboratorio lleva a cabo 3 análisis repetidos del comprimido. El resultado medio es $\bar{x} = 0{,}8404$. La verdadera concentración es la media μ de la población constituida por todos los análisis del comprimido. La desviación típica del procedimiento de análisis se sabe que es $\sigma = 0{,}0068$. ¿Existe evidencia significativa, a un nivel de significación del 1%, de que $\mu \neq 0{,}86$?

Paso 1: Hipótesis. Las hipótesis son:

$$H_0 : \mu = 0{,}86$$
$$H_a : \mu \neq 0{,}86$$

Paso 2: Estadístico de contraste. El estadístico z es:

$$z = \frac{0.8404 - 0.86}{0.0068/\sqrt{3}} = -4,99$$

Paso 3: Significación. Debido a que la prueba es de dos colas, comparamos $|z| = 4,99$ con el valor crítico $\alpha/2 = 0.005$ de la tabla C. Este valor crítico es $z^* = 2.576$. La figura 5.15 ilustra los valores z que son estadísticamente significativos. Debido a que $|z| > 2.576$, rechazamos la hipótesis nula y concluimos (a un nivel de significación del 1%) que la concentración no es la que se afirmaba que era. ■

PRUEBAS z CON UN NIVEL DE SIGNIFICACIÓN PREDETERMINADO PARA UNA MEDIA POBLACIONAL

Para contrastar la hipótesis de que H_0: $\mu = \mu_0$ a partir de una muestra aleatoria simple de tamaño n de una población de media desconocida μ y desviación típica conocida σ, calcula el estadístico z

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

Rechaza H_0 a un nivel de significación α en contra de la alternativa de una cola

$$H_a : \mu > \mu_0, \text{ si } z \geq z^*$$
$$H_a : \mu < \mu_0, \text{ si } z \leq -z^*$$

donde z^* es el valor crítico superior de α obtenido de la tabla C. Rechaza H_0 a un nivel de significación α en contra de una alternativa de dos colas

$$H_a : \mu \neq \mu_0, \text{ si } |z| \geq z^*$$

donde z^* es el valor crítico superior de $\alpha/2$ obtenido de la tabla C.

El resultado observado en el ejemplo 5.14 era $z = -4,99$. La conclusión de que este resultado es significativo a un nivel del 1% no da toda la información. La z observada queda mucho más allá que el valor crítico del 1%, y la evidencia en contra de H_0 es mucho más fuerte que la que sugiere un nivel de significación del 1%. El valor P

$$2P(Z \geq 4,99) = 0,0000006$$

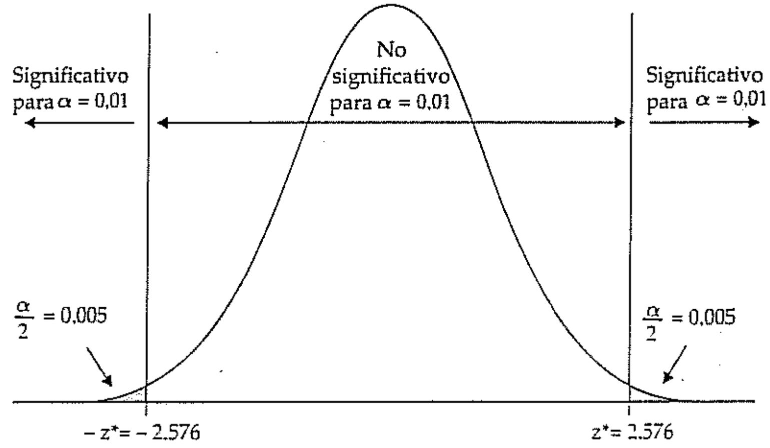


Figura 5.15. La decisión de si el estadístico z es significativo a un nivel de significación $\alpha = 0.01$ en la prueba de dos colas del ejemplo 5.14.

da una mejor información sobre la fuerza de la evidencia. *El valor P es el menor nivel α para el cual los datos son significativos.* Conocer el valor P nos permite valorar la significación a cualquier nivel.

Las tablas de los valores críticos, como la tabla C, nos permiten estimar los valores P sin realizar cálculos de probabilidad. En el ejemplo 5.14 se compara el valor observado $z = -4.99$ con *todos* los valores críticos normales de la última fila de la tabla. El valor $z = -4,99$ está más allá incluso que 3.291, el valor crítico para $P = 0,0005$. Por tanto, sabemos que para una prueba de dos colas, $P < 0.001$. En el ejemplo 5.13, $z = -1,45$ se encuentra entre los valores 0,05 y 0,10 de la tabla. Por tanto, el valor P para la prueba de una cola se encuentra entre 0.05 y 0.10. Esta aproximación es suficientemente exacta para la mayoría de los propósitos.

Debido a que en estadística aplicada casi siempre se utilizan programas estadísticos que calculan automáticamente los valores P, la utilización de la tablas de valores críticos está quedando desfasada. Las tablas de valores críticos de uso más frecuente (tales como la tabla C) aparecen en este libro con un objetivo pedagógico y como ayuda para los estudiantes que no dispongan de ordenador.

EJERCICIOS

5.40. Un ordenador tiene un generador de números aleatorios diseñado para que éstos se distribuyan uniformemente en el intervalo que va de 0 a 1. Si esto es cierto, los números generados proceden de una población con $\mu = 0{,}5$ y $\sigma = 0{,}2887$. Una instrucción para generar 100 números aleatorios da un resultado con una media $\bar{x} = 0{,}4365$. Supón que la σ de la población se mantiene fija. Queremos contrastar

$$H_0 : \mu = 0{,}5$$
$$H_a : \mu \neq 0{,}5$$

(a) Calcula el valor del estadístico z de contraste.
(b) El resultado, ¿es significativo a un nivel del 5% ($\alpha = 0{,}05$)?
(c) El resultado, ¿es significativo a un nivel del 1% ($\alpha = 0{,}01$)?

5.41. Para determinar si el contenido medio de nicotina de una marca de cigarrillos es mayor que el valor anunciado de 1,4 miligramos, se contrasta

$$H_0 : \mu = 1{,}4$$
$$H_a : \mu > 1{,}4$$

El valor calculado del estadístico de contraste es $z = 2{,}42$.
(a) El resultado, ¿es significativo a un nivel del 5%?
(b) El resultado, ¿es significativo a un nivel del 1%?

5.3.7 Pruebas derivadas de los intervalos de confianza

Los cálculos del ejemplo 5.14 para una prueba de significación del 1% son muy similares a los del ejemplo 5.4 para un intervalo de confianza del 99%. De hecho, una prueba de significación de dos colas de nivel α se puede llevar a cabo directamente a partir de un intervalo de confianza con un nivel de confianza $C = 1 - \alpha$.

INTERVALOS DE CONFIANZA Y PRUEBAS DE DOS COLAS

Una prueba de significación de dos colas de nivel α rechaza la hipótesis H_0: $\mu = \mu_0$ exactamente cuando el valor de μ_0 se sitúa fuera de un intervalo de confianza de nivel $1 - \alpha$ para μ.

EJEMPLO 5.15

El intervalo de confianza del 99% para μ del ejemplo 5.4 es:

$$\bar{x} \pm z^* \frac{\sigma}{\sqrt{n}} = 0{,}8404 \pm 0{,}0101 =$$
$$= (0{,}8303, 0{,}8505)$$

El valor hipotético $\mu = 0{,}86$ del ejemplo 5.14 se sitúa fuera de este intervalo de confianza, por lo cual rechazamos

$$H_0 : \mu = 0.86$$

a un nivel de significación del 1%. Por otro lado, no podemos rechazar

$$H_0 : \mu = 0{,}85$$

a un nivel de significación del 1% a favor de una alternativa de dos colas H_a: $\mu \neq 0{,}85$, ya que 0,85 se encuentra dentro del intervalo de confianza del 99% para μ. La figura 5.16 ilustra estos dos casos. ■

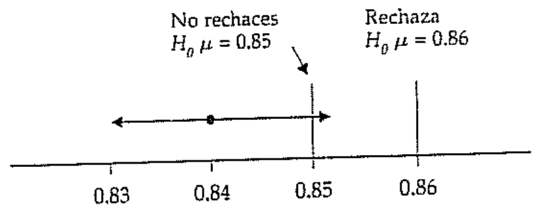


Figura 5.16. Los valores de μ que se encuentran fuera del intervalo de confianza del 99% se pueden rechazar a un nivel de significación del 1%. Los valores que caen dentro del intervalo no se pueden rechazar.

EJERCICIOS

5.42. El radón es un gas incoloro e inodoro que liberan de forma natural las rocas y suelos, y que se puede concentrar en las casas poco ventiladas. Debido a que el radón es ligeramente radiactivo, existe una cierta inquietud entre la gente por si pudiera constituir un riesgo para la salud. Los que lo desean pueden comprar detectores de radón, aunque no suelen ser muy precisos. Los investigadores de una universidad colocaron 12 detectores en una cámara donde fueron expuestos a 105 picocuries por litro de radón durante 3 días. Aquí tienes las lecturas que dieron los detectores.[7]

[7] Datos proporcionados por Diana Schellenberg, Departamento de Ciencias de la Salud, Purdue University.

91,9	97,8	111,4	122,3	105,4	95,0
103,8	99,6	96,6	119,3	104,8	101,7

Supón (aunque sea poco realista) que sabes que la desviación típica de las lecturas de los detectores de este tipo es $\sigma = 9$.

(a) Da un intervalo de confianza del 90% para la media de las lecturas μ con este tipo de detectores.

(b) ¿Existe evidencia significativa a un nivel del 10% de que la lectura media difiere del verdadero valor 105? Plantea las hipótesis y haz una prueba basándote en el intervalo de confianza del ejercicio (a).

RESUMEN

Las **pruebas de significación** se proponen valorar la evidencia proporcionada por los datos en contra de una **hipótesis nula** H_0 y a favor de una **hipótesis alternativa** H_a.

Las hipótesis se expresan en términos de parámetros poblacionales. Normalmente H_0 afirma la ausencia de efectos, y H_a establece que un determinado parámetro difiere del valor que le otorga la hipótesis nula en una dirección concreta (**pruebas de una cola**) o en cualquier dirección (**pruebas de dos colas**).

Los razonamientos esenciales de una prueba de significación son los siguientes. Supón que la hipótesis nula sea cierta. ¿Si repitiéramos la obtención de los datos muchas veces, obtendríamos a menudo datos tan inconsistentes con H_0 como los que realmente tenemos? Si los datos son poco probables cuando H_0 es cierta, son una evidencia en contra de H_0.

Una prueba de significación se basa en un **estadístico de contraste**. El **valor** P es la probabilidad, calculada suponiendo que H_0 sea cierta, de que el estadístico de contraste tome un valor al menos tan extremo como el valor observado. Los valores P pequeños indican la existencia de una fuerte evidencia en contra de H_0. El cálculo de los valores P exige conocer la distribución muestral del estadístico de contraste cuando H_0 es cierta.

Si el valor P es pequeño, o más pequeño que un valor concreto α, los datos son **estadísticamente significativos** a un nivel de significación α.

Las pruebas de significación para la hipótesis H_0: $\mu = \mu_0$, referidas al parámetro desconocido μ de una población, se basan en el **estadístico** z

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

La prueba z supone que los datos son una muestra aleatoria simple de tamaño n, que la desviación típica poblacional σ es conocida y que la población muestreada es

normal o bien que la muestra es grande. Los valores P se calculan a partir de la distribución normal estandarizada (tabla A). En las pruebas con un nivel de significación predeterminado α, se utilizan los **valores críticos** de la tabla normal estandarizada (la última fila de la tabla C).

EJERCICIOS DE LA SECCIÓN 5.3

5.43. El estudio sobre la satisfacción en el trabajo del ejemplo 5.8 determinó el grado de satisfacción en el trabajo de 28 mujeres en una cadena de montaje trabajando a su propio ritmo y, alternativamente, al ritmo fijado por una máquina. El parámetro μ es la media de las diferencias entre ambos resultados en la población de este tipo de trabajadoras. Los resultados están distribuidos normalmente. La desviación típica de la población es $\sigma = 0{,}60$. Las hipótesis son

$$H_0 : \mu = 0$$
$$H_a : \mu \neq 0$$

(a) ¿Cuál es la distribución de la media de los resultados $\bar{x}$ del estudio de la satisfacción en el trabajo de las 28 trabajadoras, si la hipótesis nula es cierta? Dibuja la curva de densidad de esta distribución. (Sugerencia: dibuja primero una curva normal, luego señala en el eje de las abcisas lo que sabes sobre la localización de μ y σ en una curva normal).

(b) Supón que el estudio halló que $\bar{x} = 0{,}09$. Señala este punto en el eje de las abcisas de tu dibujo. En realidad, el estudio halló que para estas 28 trabajadoras, $\bar{x} = 0{,}27$. Señala este punto en tu dibujo. Basándote en tu gráfico, explica con un lenguaje sencillo por qué uno de los resultados constituye una buena evidencia de que H_0 no es cierta y el otro no.

(c) Dibuja otra vez la curva normal. Sombrea el área por debajo de la curva que da el valor P para el resultado $\bar{x} = 0{,}09$. Luego calcula este valor P (fíjate en que H_a es de dos colas).

(d) Calcula también el valor P para el resultado $\bar{x} = 0{,}27$. Los dos valores P expresan tu explicación de (b) de forma numérica.

5.44. Se anuncia que la superficie media de varios miles de apartamentos de un determinado complejo urbanístico es de 116 metros cuadrados. Un grupo de inquilinos cree que los apartamentos son más pequeños de lo que se anuncia. En consecuencia, deciden contratar a un aparejador para que determine la superficie de una muestra de apartamentos con el fin de comprobar su sospecha. ¿Cuáles son las hipótesis nula H_0 y alternativa H_a?

do cuidadosamente la forma de identificarlo. En otros casos, las pruebas de significación pueden tener poco sentido.

EJEMPLO 5.18

Supongamos que quieres saber qué es lo que distingue a los ejecutivos en formación que llegan a ocupar cargos de responsabilidad en la empresa de los que acaban abandonándola. Tienes muchos datos de todos los ejecutivos en formación que han pasado por la empresa (datos sobre su personalidad, sobre sus objetivos, sobre su formación universitaria, incluso sobre su familia y sus aficiones). Los programas estadísticos nos permiten hacer, sin la menor dificultad, docenas de pruebas de significación sobre todas estas variables para ver cuáles predicen mejor el éxito final de los ejecutivos. ¡Aja!, descubres que los futuros ejecutivos tienen significativamente más probabilidades de haberse criado en un entorno urbano, y tener un título universitario en una carrera técnica que los que terminan marchándose de la empresa.

Antes de recomendar que la contratación futura se haga teniendo en cuenta estos hallazgos, deja pasar un poco de tiempo y reflexiona. Cuando haces docenas de pruebas de significación a un nivel del 5%, cabe esperar que algunas de estas pruebas sean significativas sólo por azar. Después de todo, los resultados significativos a un nivel del 5% ocurren por azar 5 veces de cada 100, después de muchas repeticiones, incluso cuando H_0 es cierta. Llevar a cabo una prueba y alcanzar un nivel de significación $\alpha = 0.05$, es una garantía razonable de que has hallado algún efecto. Hacer docenas de pruebas y alcanzar este nivel de significación una vez o dos no lo es. ■

Existen métodos para contrastar muchas hipótesis simultáneamente que permiten controlar el riesgo de hallar significaciones equivocadas. Pero si llevas a cabo muchas pruebas individuales sin utilizar estos métodos especiales, hallar algunos valores P pequeños constituye sólo un indicio, pero no un resultado concluyente. Se puede decir lo mismo de los análisis menos formales. Buscar la variable para la cual son mayores las diferencias entre el grupo de jóvenes ejecutivos que llegan a ocupar cargos de responsabilidad y el grupo de los que finalmente no continúan, y luego hacer una prueba de significación para contrastar si las diferencias son significativas es una mala utilización de la estadística. El valor P supone que tu ya sabías qué comparaciones querías hacer antes de observar los datos. El valor P es muy engañoso cuando se aplica a la mayor de muchas diferencias.

Explorar los datos para hallar regularidades o irregularidades es sin duda legítimo. El análisis exploratorio de datos es un aspecto importante de la estadística. Pero los razonamientos de la inferencia estadística no se pueden aplicar cuando consigues encontrar en los datos algún efecto sorprendente. El remedio es claro. Cuando tengas

una hipótesis, diseña un estudio para buscar el efecto concreto que crees que existe. Si el resultado de este estudio es estadísticamente significativo, has encontrado una evidencia real.

EJERCICIOS

5.59. Un investigador que busca evidencias a favor de la percepción extrasensorial hace pasar una prueba a 500 sujetos. Cuatro de estos sujetos lo hacen significativamente mejor ($P < 0.01$) que responder al azar.

(a) ¿Es correcto llegar a la conclusión de que estas cuatro personas tienen percepción extrasensorial? Justifica tu respuesta.

(b) ¿Qué debería hacer ahora el investigador para verificar que cualquiera de estas cuatro personas tiene percepción extrasensorial?

RESUMEN

Los valores P aportan más información que simplemente rechazar o no el resultado de una prueba de significación con un nivel de significación α predeterminado. No des excesiva importancia a los valores tradicionales de α, tales como $\alpha = 0{,}05$.

Los efectos muy pequeños pueden ser muy significativos (valores P pequeños), especialmente cuando la prueba se basa en una muestra grande. Un efecto estadísticamente significativo no tiene por qué ser importante en la práctica. Representa gráficamente los datos para mostrar el efecto que estás buscando y utiliza los intervalos de confianza para estimar los verdaderos valores de los parámetros.

Por otro lado, la falta de significación no significa que H_0 sea cierta, especialmente cuando la prueba se basa sólo en unas pocas observaciones.

Las pruebas de significación no siempre son válidas. La obtención de datos de forma incorrecta, la presencia de observaciones atípicas o el contraste de las hipótesis que sugieren los propios datos pueden invalidar una prueba.

La realización simultánea de muchas pruebas de significación probablemente dará lugar a algunos resultados significativos sólo por azar, incluso si todas las hipótesis nulas son ciertas.

EJERCICIOS DE LA SECCIÓN 5.4

5.60. ¿A cuáles de las siguientes preguntas responde una prueba de significación?

(a) ¿Se ha diseñado correctamente la muestra o el experimento?

(b) El efecto observado, ¿se debe al azar?
(c) El efecto observado, ¿es importante?

5.61. Una empresa compara dos diseños de envases de detergente para lavadoras mediante la colocación de botes con ambos diseños en los estantes de varias tiendas. Los datos sobre más de 5.000 botes comprados indican que el Diseño A fue comprado por más clientes que el Diseño B. La diferencia es estadísticamente significativa ($P = 0{,}02$). ¿Podemos afirmar que los consumidores prefieren claramente el Diseño A al Diseño B? Justifica tu respuesta.

5.62. Una vez, un grupo de psicólogos midió 77 variables de una muestra de personas esquizofrénicas y de una muestra de personas que no lo eran. Los psicólogos compararon las dos muestras utilizando 77 pruebas de significación distintas. Dos de estas pruebas fueron significativas a un nivel del 5%. Supón que en realidad no hay diferencias entre las dos poblaciones en ninguna de las 77 variables. Por tanto, las 77 hipótesis nulas son ciertas.

(a) ¿Cuál es la probabilidad de que una prueba determinada dé una diferencia significativa a un nivel del 5%?

(b) ¿Por qué no es sorprendente que 2 de las 77 pruebas fueran significativas a un nivel del 5%?

5.5 La inferencia como método para tomar decisiones*

Las pruebas de significación valoran la fuerza de la evidencia en contra de la hipótesis nula. La fuerza de la evidencia en contra de H_0 la medimos mediante el valor P, que es una probabilidad calculada bajo el supuesto de que H_0 sea cierta. La hipótesis alternativa H_a (la afirmación para la que buscamos evidencia a favor) interviene en la prueba sólo para ayudarnos a determinar qué resultados cuentan en contra de H_0.

De todas formas, la utilización de las pruebas con un nivel de significación α predeterminado sugieren otra cosa. Un nivel de significación α escogido antes de hacer la prueba deja entrever que los resultados de la prueba se utilizarán para tomar una *decisión*. Si nuestro resultado es significativo a un nivel α, rechazamos H_0 en favor de H_a. En caso contrario, no podemos rechazar H_0. El paso desde medir la fuerza de la evidencia en contra de H_0 hasta tomar una decisión no es pequeño. Muchos estadísticos creen que la responsabilidad de tomar una decisión se debe

* El objetivo de esta sección más avanzada es clarificar el sentido de las pruebas de significación al compararlas con otros métodos similares. Esta sección no es necesaria para leer el resto del libro.

dejar al usuario y que no tiene que formar parte de la prueba. Los resultados de una prueba son únicamente uno más entre los muchos factores que influyen en una decisión.

Controles de calidad

De todas formas, hay circunstancias que exigen tomar decisiones después de la inferencia. Los ***controles de calidad*** constituyen una de estas circunstancias. Un fabricante de cojinetes y la empresa compradora de éstos llegan a un acuerdo sobre unos requisitos específicos de calidad que deben cumplir los envíos de cojinetes. Cuando llega un envío, la empresa compradora inspecciona una muestra de los cojinetes. En función del resultado del muestreo, la empresa compradora acepta o no el envío. Utilizaremos los controles de calidad para mostrar cómo un concepto distinto –la inferencia como decisión– cambia la argumentación utilizada hasta ahora en las pruebas de significación.

5.5.1 Errores de tipo I y errores de tipo II

Las pruebas de significación se concentran en H_0, la hipótesis nula. Sin embargo, si hay que tomar una decisión, no hay nada que justifique que nos fijemos sólo en H_0. Simplemente hay dos hipótesis, y tenemos que aceptar una y rechazar la otra. Es conveniente que sigamos llamando a las dos hipótesis H_0 y H_a, pero a partir de ahora H_0 (la afirmación en contra de la cual queremos encontrar evidencia) ya no tendrá el papel destacado que tenía en las pruebas de significación. En los problemas de control de calidad tenemos que decidir entre

H_0: El envío de cojinetes cumple los estándares de calidad.

H_a: El envío de cojinetes no cumple los estándares de calidad,

a partir de una muestra de cojinetes.

Confiamos en que nuestra decisión será la correcta, aunque algunas veces nos equivocaremos. Hay dos tipos de decisiones incorrectas: aceptar un envío de cojinetes defectuosos y rechazar un envío de cojinetes buenos. Aceptar un envío defectuoso perjudica al consumidor, mientras que rechazar un envío bueno perjudica al vendedor. Para distinguir estos dos tipos de errores, los llamamos de maneras distintas.

ERRORES DE TIPO I Y ERRORES DE TIPO II

Si rechazamos H_0 (aceptamos H_a) cuando en realidad H_0 es cierta, cometemos un error de tipo I.

Si aceptamos H_0 (rechazamos H_a) cuando en realidad H_a es cierta, cometemos un error de tipo II.

La figura 5.17 muestra las cuatro situaciones posibles. Cuando H_0 es cierta, nuestra decisión es correcta (si aceptamos H_0) o cometemos un error de tipo I. Cuando H_a es cierta, nuestra decisión es correcta o cometemos un error de tipo II. En cada ocasión sólo es posible cometer un tipo de error.

		Certeza sobre la población	
		H_0 cierta	H_a cierta
Decisión basada en la muestra	Rechazo de H_0	Error de tipo I	Decisión correcta
	Aceptación de H_0	Decisión correcta	Error de tipo II

Figura 5.17. Los dos tipos de error en el contraste de hipótesis.

5.5.2 Probabilidades de error

Valoraremos cualquier procedimiento de toma de decisiones examinando las probabilidades de los dos tipos de error. De esta forma nos mantenemos en el principio de que la inferencia estadística se basa en preguntarse, "¿qué ocurriría si utilizáramos este procedimiento muchas veces?"

Las pruebas de significación con un nivel α predeterminado proporcionan una guía para tomar decisiones, porque la prueba, o bien rechaza H_0 o bien no la consigue rechazar. Desde la perspectiva de quienes tienen que tomar decisiones, no conseguir rechazar H_0 significa decidir que H_0 es cierta. Podemos, pues, describir el

resultado de una prueba mediante las probabilidades de los errores de tipo I y de los errores de tipo II.

EJEMPLO 5.19

El diámetro medio de un determinado tipo de cojinetes se supone que es de 2.000 centímetros (cm). El diámetro de los cojinetes varía según una distribución normal con una desviación típica σ = 0,010 cm. Cuando llega un envío de cojinetes, el comprador toma una muestra aleatoria simple de 5 cojinetes del lote y mide sus diámetros. El comprador rechaza el lote si el diámetro medio de la muestra es significativamente distinto de 2 cm a un nivel de significación del 5%.

Lo que hace el comprador es un contraste de las hipótesis

$$H_0 : \mu = 2$$
$$H_a : \mu \neq 2$$

Para llevar a cabo la prueba, el comprador calcula el estadístico z

$$z = \frac{\bar{x} - 2}{0.01/\sqrt{5}}$$

y rechaza H_0 si $z < -1.96$ o si $z > 1.96$. Cometer un error de tipo I significa rechazar H_0 cuando en realidad $\mu = 2$.

¿Qué podemos decir sobre los errores de tipo II? Puesto que H_a admite muchos valores posibles de μ, nos centraremos en uno de estos valores. El comprador y el vendedor llegan a un acuerdo que consiste en que se deberá rechazar un envío de cojinetes cuando su diámetro medio sea igual a 2.015 cm. Por tanto, se cometerá un error de tipo II cuando se acepte H_0 siendo en realidad $\mu = 2{,}015$.

La figura 5.18 muestra cómo se obtienen las dos probabilidades de error a partir de las *dos* distribuciones muestrales de $\bar{x}$, para $\mu = 2$ y para $\mu = 2.015$. Cuando $\mu = 2$, H_0 es cierta y rechazar H_0 significa cometer un error de tipo I. Cuando $\mu = 2{,}015$, H_a es cierta y aceptar H_0 significa cometer un error de tipo II. A continuación calcularemos las probabilidades de cada uno de estos tipos de error. ■

La probabilidad de un error de tipo I es la probabilidad de rechazar H_0 cuando en realidad H_0 es cierta. Ésta es la probabilidad de que $|z| \geq 1{,}96$, cuando $\mu = 2$. Pero éste es exactamente el nivel de significación de la prueba. El valor crítico 1,96 se escogió para hacer que esta probabilidad fuera 0,05, por lo cual no la tenemos que calcular otra vez. La definición de "nivel de significación de 0,05" significa que

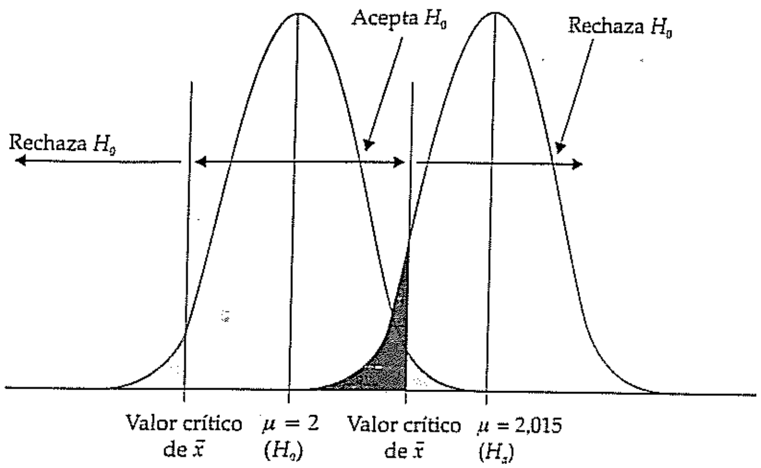


Figura 5.18. Las probabilidades de los dos tipos de error del ejemplo 5.19. La probabilidad del error de tipo I (*el área sombreada más clara*) es la probabilidad de rechazar $H_0: \mu = 2$ cuando en realidad $\mu = 2$. La probabilidad de un error de tipo II (*el área sombreada más oscura*) es la probabilidad de aceptar H_0 cuando en realidad $\mu = 2.015$.

valores z tan extremos ocurrirán con una probabilidad igual a 0.05 cuando H_0 sea cierta.

NIVEL DE SIGNIFICACIÓN Y ERROR DE TIPO I

El nivel de significación α de cualquier prueba de significación con un nivel predeterminado es la probabilidad de un error de tipo I. Es decir, α es la probabilidad de que la prueba rechace la hipótesis nula H_0 cuando H_0 en realidad es cierta.

La probabilidad de un error de tipo II para la alternativa $\mu = 2{,}015$ del ejemplo 5.19 es la probabilidad de que la prueba acepte H_0 cuando μ toma este valor alternativo. Ésta es la probabilidad de que el estadístico z se encuentre entre $-1{,}96$ y $1{,}96$, calculado suponiendo que $\mu = 2{,}015$. Esta probabilidad *no* es $1 - 0{,}05$, ya que la probabilidad 0,05 se halló suponiendo que $\mu = 2$. He aquí el cálculo del error de tipo II.

EJEMPLO 5.20

Para calcular la probabilidad de un error de tipo II:

Paso 1. *Escribe la regla de aceptación de H_0 en términos de $\bar{x}$.* La prueba acepta H_0 cuando

$$-1{,}96 \le \frac{\bar{x} - 2}{0{,}01/\sqrt{5}} \le 1{,}96$$

que es lo mismo que,

$$2 - 1{,}96\left(\frac{0{,}01}{\sqrt{5}}\right) \le \bar{x} \le 2 + 1{,}96\left(\frac{0{,}01}{\sqrt{5}}\right)$$

o que, una vez realizado el cálculo,

$$1{,}9912 \le \bar{x} \le 2{,}0088$$

En este paso no interviene la alternativa de que $\mu = 2.015$.

Paso 2. *Halla la probabilidad de aceptar H_0 suponiendo que la hipótesis alternativa sea cierta.* Toma $\mu = 2.015$ y estandariza para hallar la probabilidad.

$$\begin{aligned} P(\text{error de tipo II}) &= P(1{,}9912 \le \bar{x} \le 2.0088) \\ &= P\left(\frac{1.9912 - 2.015}{0{,}01/\sqrt{5}} \le \frac{\bar{x} - 2.015}{0.01/\sqrt{5}} \le \frac{2.0088 - 2.015}{0.01/\sqrt{5}}\right) = \\ &= P(-5.32 \le Z \le -1.39) = 0{,}0823 \end{aligned}$$

La figura 5.19 ilustra esta probabilidad de error en términos de la distribución de $\bar{x}$ cuando $\mu = 2{,}015$. La prueba aceptaría de una manera equivocada la hipótesis de que $\mu = 2$ en aproximadamente un 8% de todas las muestras cuando $\mu = 2{,}015$. ■

Esta prueba de significación rechazará un 5% de todos los envíos de cojinetes correctos (para los cuales $\mu = 2$). Aceptará un 8% de los envíos de forma equivocada cuando $\mu = 2{,}015$. Los cálculos de las probabilidades de los errores ayudan al vendedor y al comprador a decidir si la prueba es satisfactoria.

EJERCICIOS

5.63. Tu empresa comercializa un programa de diagnóstico médico informatizado. El programa comprueba los resultados de las pruebas médicas rutinarias (presión san-

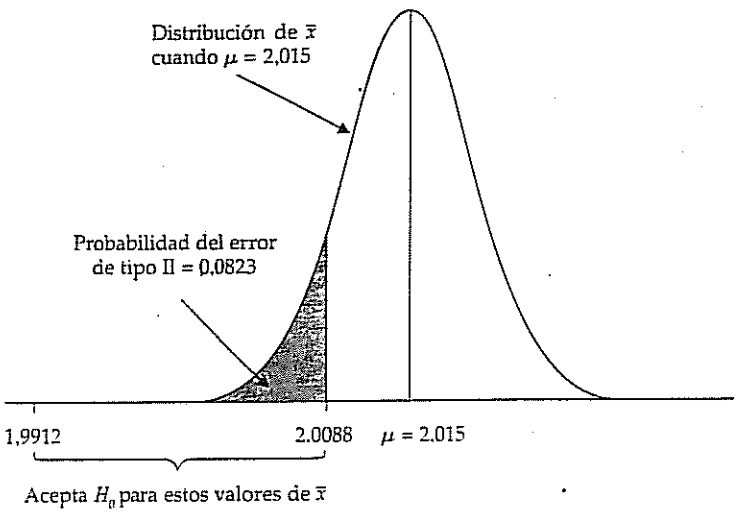


Figura 5.19. La probabilidad del error de tipo II para el ejemplo 5.20. Ésta es la probabilidad de que la prueba acepte H_0 cuando la hipótesis alternativa es cierta.

guínea, análisis de sangre, etc.). El programa se utiliza para filtrar a miles de personas en cuyos análisis no se detecta ninguna anomalía. En cada caso, el programa toma una decisión.

(a) ¿Cuáles son las dos hipótesis y los dos tipos de error que el programa puede cometer? Describe los dos tipos de error en términos de "falsos positivos" y de "falsos negativos".

(b) El programa se puede ajustar para disminuir la probabilidad de un tipo de error a costa de aumentar la probabilidad del otro tipo. ¿Qué probabilidad de error harías más pequeña y por qué? (Esta decisión es subjetiva. No existe una sola respuesta correcta).

5.64. Tienes los resultados de la prueba aritmética de la encuesta NAEP para una muestra aleatoria simple de 840 hombres jóvenes. Quieres contrastar las siguientes hipótesis sobre la media de los resultados de la población.

$$H_0 : \mu = 275$$
$$H_a : \mu < 275$$

a un nivel de significación del 1%. La desviación típica poblacional se sabe que es $\sigma = 60$. El estadístico z de contraste es

$$z = \frac{\bar{x} - 275}{60/\sqrt{840}}$$

(a) ¿Qué procedimiento se sigue para rechazar H_0 en términos de z?

(b) ¿Cuál es la probabilidad de un error de tipo I?

(c) Quieres saber si esta prueba rechazará a menudo H_0 cuando la verdadera media poblacional sea 270, es decir, 5 unidades menor de lo que afirma la hipótesis nula. Contesta esta pregunta calculando la probabilidad de un error de tipo II cuando $\mu = 270$.

5.65. Tienes una muestra aleatoria simple de tamaño $n = 9$ de una distribución normal con $\sigma = 1$. Deseas contrastar

$$H_0 : \mu = 0$$
$$H_a : \mu > 0$$

Decides rechazar H_0 si $\bar{x} > 0$ y aceptar H_0 en cualquier otro caso.

(a) Halla la probabilidad de un error de tipo I. Es decir, halla la probabilidad de que la prueba rechace H_0 cuando en realidad $\mu = 0$.

(b) Halla la probabilidad de un error de tipo II cuando $\mu = 0.3$. Es decir, halla la probabilidad de que la prueba acepte H_0 cuando en realidad $\mu = 0.3$.

(c) Halla la probabilidad de un error de tipo II cuando $\mu = 1$.

5.5.3 Potencia

Una prueba comete un error de tipo II cuando no rechaza una hipótesis nula que en realidad es falsa. Una probabilidad alta de un error de tipo II para una determinada alternativa, significa que la prueba no es, a menudo, suficientemente sensible para detectar la alternativa. Los cálculos de la probabilidad de los errores de tipo II son, por tanto, útiles incluso si no piensas que una prueba estadística se puede considerar como un método para tomar decisiones. El lenguaje que se utiliza cuando se

POTENCIA

La probabilidad de que una prueba, con un nivel de significación predeterminado α, rechace H_0 cuando un cierto valor del parámetro alternativo es verdadero, se llama **potencia** de la prueba en contra de esa alternativa.

La potencia de una prueba en contra de cualquier alternativa es igual a 1 menos la probabilidad de un error de tipo II para esa alternativa.

toman decisiones es algo distinto del lenguaje utilizado en las pruebas de significación. Cuando se toman decisiones se menciona la probabilidad de que una prueba *rechace* H_0 cuando una determinada alternativa es cierta. Cuanto más alta sea esta probabilidad, más sensible es la prueba.

Los cálculos de la potencia son esencialmente los mismos que los cálculos de probabilidad de los errores de tipo II. En el ejemplo 5.20 la potencia es la probabilidad de *rechazar* H_0 en el segundo paso del cálculo. Esta probabilidad es igual a 1 – 0,0823, o 0,9177.

Tanto el cálculo de los valores *P* como el cálculo de la potencia nos dicen lo que ocurriría si repitiéramos la prueba muchas veces. El valor *P* describe lo que ocurriría si supusiéramos que la hipótesis nula es cierta. La potencia describe lo que ocurriría si supusiéramos que una determinada alternativa es cierta.

En la preparación de una investigación que incluya pruebas de significación, un usuario prudente de la estadística decide qué alternativas debe detectar la prueba y comprueba que la potencia sea la adecuada. La potencia depende del valor concreto del parámetro en H_a en el que estemos interesados. Los valores de la media μ que están en H_a pero que se hallan próximos al hipotético valor μ_0 son más difíciles de detectar (la potencia es baja) que los valores de μ alejados de μ_0. Si la potencia es demasiado baja, una muestra mayor aumentará la potencia para un mismo nivel de significación α. Para calcular la potencia, debemos fijar un valor de α de manera que tengamos una regla fija para rechazar H_0. De todas formas, es preferible dar los valores *P* a utilizar un nivel de significación predeterminado. La práctica habitual consiste en calcular la potencia a un determinado nivel de significación como α = 0,05, incluso si se tiene la intención de dar el valor *P*.

EJERCICIOS

5.66. El fabricante de refrescos del ejercicio 5.7 considera que una pérdida de dulzura sería inaceptable si la media de las respuestas de todos los catadores es μ = 1,1. ¿Una prueba de significación del 5% para contrastar las hipótesis

$$H_0 : \mu = 0$$
$$H_a : \mu > 0$$

basada en una muestra de 10 catadores detectará por lo general un cambio de esta magnitud?

Queremos la potencia de la prueba en contra de la alternativa μ = 1,1. Esta potencia es la probabilidad de que la prueba rechace H_0 cuando μ = 1,1 es cierta. El método de cálculo es similar al método de cálculo del error de tipo II.

(a) **Paso 1:** *Escribe el procedimiento para rechazar H_0 en términos de $\bar{x}$.* Sabemos que σ = 1, por lo que la prueba rechaza H_0 a un nivel α = 0,05 cuando

$$z = \frac{\bar{x} - 0}{1/\sqrt{10}} \geq 1.645$$

Expresa esta desigualdad en términos de $\bar{x}$.

(b) **Paso 2:** *La potencia es la probabilidad de este suceso suponiendo que la alternativa sea cierta.* Estandariza la desigualdad utilizando μ = 1,1 para hallar la probabilidad de que $\bar{x}$ tome un valor que lleve al rechazo de H_0.

5.67. El ejercicio 5.39 hace referencia a una prueba sobre el contenido medio de las botellas de refrescos. Las hipótesis son

$$H_0 : \mu = 300$$
$$H_a : \mu < 300$$

El tamaño de la muestra es n = 6, y se supone que la población tiene una distribución normal con σ = 3. Una prueba de significación del 5% rechaza H_0 si $z \leq -1.645$, donde el estadístico z de contraste es

$$z = \frac{\bar{x} - 300}{3/\sqrt{6}}$$

Los cálculos de la potencia nos ayudan a determinar cuál es la magnitud del déficit en el contenido de las botellas que se espera que pueda detectar la prueba.

(a) Halla la potencia de esta prueba en contra de la alternativa μ = 299.

(b) Halla la potencia de la prueba en contra de la alternativa μ = 295.

(c) La potencia de la prueba en contra de μ = 290, ¿es mayor o menor que el valor que hallaste en (b)? (No calcules esta potencia). Justifica tu respuesta.

5.68. Aumentando el tamaño de la muestra se incrementa la potencia de una prueba cuando el nivel α no cambia. Supón que en el ejercicio anterior se hubieran tomado medidas de una muestra de n botellas. En ese ejercicio, n = 6. La prueba de significación del 5% sigue rechazando H_0 cuando $z \leq -1,645$, pero ahora el estadístico z es

$$z = \frac{\bar{x} - 300}{3/\sqrt{n}}$$

(a) Halla la potencia de esta prueba en contra de la alternativa $\mu = 299$ cuando $n = 25$.

(b) Halla la potencia de la prueba en contra de $\mu = 299$ cuando $n = 100$.

5.5.4 Diferentes puntos de vista sobre las pruebas estadísticas

La distinción entre pruebas de significación y pruebas como reglas, para decidir entre dos hipótesis, no se halla en los cálculos, sino en los razonamientos que motivan dichos cálculos. En una prueba de significación nos concentramos en una sola hipótesis (H_0) y en una sola probabilidad (el valor P). El objetivo es medir la fuerza de la evidencia muestral en contra de H_0. Los cálculos de la potencia se hacen para comprobar la sensibilidad de la prueba. Si no podemos rechazar H_0, sólo estamos diciendo que no existe suficiente evidencia en contra de H_0, no que H_0 sea realmente cierta. Si el mismo problema de inferencia se considera como un problema de toma de decisiones, nos fijamos en las dos hipótesis por igual y establecemos una regla para decidir entre ellas a partir de la evidencia muestral. Es decir, nos concentramos en dos probabilidades, las probabilidades de los dos tipos de error. Debemos escoger una de las dos hipótesis y no podemos rehuir la decisión aduciendo que no hay suficiente evidencia.

Contraste de hipótesis

Existen diferencias claras entre las dos maneras de considerar las pruebas estadísticas. De tomas formas, algunas veces ambos puntos de vista convergen. Jerzy Neyman defendió una versión llamada ***contraste de hipótesis*** que mezcla la idea de prueba de significación y la idea de regla para tomar decisiones de la siguiente manera:

1. Plantea H_0 y H_a como si se tratase de una prueba de significación. Es decir, estamos buscando una evidencia en contra de H_0.
2. Piensa en el problema como si se tratara de un problema de toma de decisiones, por lo cual las probabilidades de los errores de tipo I y de tipo II pasan a ser relevantes.
3. Por lo que decimos en 1, los errores de tipo I son más importantes. Por tanto, escoge un α (el nivel de significación) y considera sólo las pruebas con una probabilidad de error de tipo I que no sea mayor que α.
4. Entre todas estas pruebas escoge la que tenga una probabilidad de error de tipo II tan pequeña como sea posible (es decir, que la potencia sea la mayor posible). Si esta probabilidad es demasiado grande, tendrás que tomar una muestra mayor para reducir las posibilidades de error.

Al contraste de hipótesis se le da, a menudo, mucho énfasis en los textos de estadística matemática, debido a que Neyman desarrolló una teoría matemática impresionante. En casos fáciles, esta teoría muestra cómo hallar la prueba que tiene la probabilidad de error de tipo II más pequeña entre todas las pruebas que tienen una determinada probabilidad (como por ejemplo 0,05) de cometer un error de tipo I. En parte, debido a que resultados de este tipo no son posibles en muchas situaciones prácticas, la idea que subyace en las pruebas de significación es la que prevalece en la estadística aplicada.

RESUMEN

Una alternativa a las pruebas de significación considera H_0 y H_a como dos hipótesis de igual importancia entre las que tenemos que decidir. Este punto de vista propio del análisis de decisiones considera, en general, la inferencia estadística como un método para tomar decisiones en presencia de la incertidumbre.

En el caso de que se contraste H_0 en contra de H_a, el análisis de decisiones escoge una regla de decisión basada en las probabilidades de los dos tipos de error. Cometemos un error de tipo I si rechazamos H_0 cuando en realidad H_0 es cierta. Cometemos un error de tipo II si aceptamos H_0 cuando en realidad H_a es cierta.

La potencia de una prueba de significación mide la capacidad de la prueba para detectar una hipótesis alternativa. La potencia en contra de una determinada alternativa es la probabilidad de que la prueba rechace H_0 cuando la alternativa sea cierta.

En una prueba de significación con un nivel de significación α predeterminado, el nivel de significación α es la probabilidad del error de tipo I, y la potencia en contra de una determinada alternativa es igual a 1 menos la probabilidad del error de tipo II de esa alternativa.

Un aumento del tamaño de la muestra incrementa la potencia (reduce la probabilidad del error de tipo II) cuando el nivel de significación se mantiene fijo.

EJERCICIOS DE LA SECCIÓN 5.5

5.69. Los cálculos de la potencia de las pruebas de dos colas siguen la misma pauta que los cálculos de la potencia de las pruebas de una cola. El ejemplo 5.14 presenta el contraste de dos colas

$$H_0 : \mu = 0{,}86$$
$$H_a : \mu \neq 0{,}86$$

a un nivel de significación del 1%. El tamaño de la muestra es $n = 3$ y $\sigma = 0{,}0068$. Hallaremos la potencia de esta prueba en contra de la alternativa $\mu = 0{,}845$.

(a) La prueba del ejemplo 5.14 rechaza H_0 cuando $|z| \geq 2{,}576$. El estadístico z de contraste es

$$z = \frac{\bar{x} - 0{,}86}{0{,}0068/\sqrt{3}}$$

Describe la regla para rechazar H_0 en términos de los valores de $\bar{x}$ (debido a que la prueba es de dos colas, se rechaza H_0 cuando $\bar{x}$ es demasiado grande o demasiado pequeña).

(b) Ahora halla la probabilidad de que $\bar{x}$ tome valores que conduzcan a rechazar H_0 si la verdadera media poblacional es $\mu = 0{,}845$. Esta probabilidad es la potencia de la prueba.

(c) ¿Cuál es la probabilidad de que esta prueba cometa un error de tipo II cuando $\mu = 0{,}845$?

5.70. En el ejemplo 5.11 el médico de una empresa no halló evidencia significativa de que la media de la presión sanguínea de una población de ejecutivos fuera distinta de la media nacional $\mu = 128$. El médico se pregunta ahora si la prueba utilizada detectaría una diferencia importante si la hubiera. Para una muestra aleatoria simple de tamaño 72 de una población con una desviación típica $\sigma = 15$, el estadístico z es

$$z = \frac{\bar{x} - 128}{15/\sqrt{72}}$$

La prueba de dos colas rechaza $H_0: \mu = 128$ a un nivel de significación del 5% cuando $|z| \geq 1{,}96$.

(a) Halla la potencia de la prueba en contra de la alternativa $\mu = 134$.

(b) Halla la potencia de la prueba en contra de $\mu = 122$. ¿Se puede confiar en que la prueba detecte una media que esté a 6 unidades de la media $\mu = 128$?

(c) Si la alternativa estuviera más lejos de H_0, digamos que $\mu = 136$, la potencia de la prueba, ¿sería mayor o menor que los valores calculados en (a) y (b)?

5.71. En el ejercicio 5.67 hallaste la potencia de una prueba en contra de la alternativa $\mu = 295$. Utiliza el resultado de ese ejercicio para hallar las probabilidades de los errores de tipo I y de tipo II para esa prueba y esa alternativa.

5.72. En el ejercicio 5.64 hallaste las probabilidades de los dos tipos de error de la prueba $H_0: \mu = 275$, con la alternativa concreta $\mu = 270$. Utiliza el resultado de ese ejercicio para dar la potencia de la prueba en contra de la alternativa $\mu = 270$.