6. 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 en 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.

6.1 Introducción

Cuando obtenemos una muestra, conocemos las respuestas de cada uno de sus individuos. No obstante, en general, no tenemos suficiente con la información de la muestra. Queremos *inferir* a partir de los datos de la muestra alguna conclusión sobre la población que ésta representa.

INFERENCIA ESTADÍSTICA

La **inferencia estadística** proporciona métodos que permiten sacar conclusiones de una población a partir de los datos de una muestra.

No podemos estar seguros de que nuestras conclusiones sean correctas —otra muestra podría conducirnos a otras conclusiones—. La inferencia estadística utiliza el lenguaje de la probabilidad para indicar la fiabilidad de sus conclusiones.

En este capítulo nos encontraremos las dos modalidades principales de inferencia estadística. La sección 6.2 trata sobre los *intervalos de confianza*, utilizados para la estimación del valor de un parámetro poblacional. La sección 6.3 introduce las *pruebas de significación*, utilizadas para valorar la evidencia de una afirmación sobre una población. Las dos modalidades de inferencia se basan en la distribución muestral de estadísticos. Es decir, las dos modalidades de inferencia proporcionan probabilidades que nos informan sobre lo que ocurriría si utilizáramos el método de inferencia muchas veces.

Los razonamientos de la inferencia estadística al igual que la probabilidad tratan sobre las regularidades que aparecen después de muchas repeticiones. La inferencia es más fiable cuando los datos se han obtenido a partir de un diseño aleatorio diseñado de forma correcta. **Cuando utilices la inferencia estadística, actúa como si tus datos fueran una muestra aleatoria o procedieran de un diseño aleatorizado**. Si no lo hicieras de esta manera, tus conclusiones pueden ser fácilmente criticadas. Las pruebas de significación y los intervalos de confianza no pueden solucionar problemas en la obtención de datos tales como las muestras de voluntarios o los experimentos incontrolados. Utiliza el análisis de datos que estudiaste durante los primeros tres capítulos de este libro y aplica la inferencia sólo cuando estés seguro de que realmente se puede utilizar con tus datos.

Este capítulo introduce los razonamientos utilizados en inferencia estadística. Ilustraremos los razonamientos con algunas técnicas de inferencia concretas.

De todas formas estas técnicas se han simplificado de tal forma que no resultan muy útiles en la práctica. En los próximos capítulos veremos cómo modificar estas técnicas de manera que resulten aplicables a casos reales. En los próximos capítulos también se introducirán métodos de inferencia utilizables en la mayoría de situaciones prácticas que encontramos cuando aprendimos cómo explorar los datos. En las bibliotecas existen libros y programas estadísticos llenos de técnicas más elaboradas. La utilización correcta de cualquiera de ellas exige que comprendas su fundamento. Un ordenador puede hacer los cálculos, pero tú debes decidir, basándote en tus conocimientos, si el método es adecuado.

6.2 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 estadounidenses 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 6.1. Resultados de la encuesta NAEP

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 gramo a dólares por kilo.

En un año reciente participaron en la encuesta NAEP 840 hombres entre 21 y 25 años. Su puntuación media 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 de estas edades?[1] ■

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

La ley de los grandes números nos dice que la media $\bar{x}$ de una gran muestra aleatoria simple tomará un valor próximo a 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 $\bar{x}$:

- El teorema del límite central nos indica que una media $\bar{x}$ calculada a partir de 840 notas 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 $\frac{\sigma}{\sqrt{840}}$, 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 $\bar{x}$ será

$$\frac{\sigma}{\sqrt{n}} = \frac{60}{\sqrt{840}} \doteq 2{,}1$$

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

Si seleccionáramos muchas muestras repetidas de tamaño 840 y hallásemos la puntuación media de cada una de ellas, podríamos obtener la media $\bar{x} = 272$ de la primera muestra, $\bar{x} = 268$ de la segunda muestra, $\bar{x} = 273$ de la tercera muestra, etc. Si representáramos de forma gráfica la distribución de estas medias, obtendríamos la distribución normal con media igual a la media desconocida de la población y desviación típica igual a 2,1. La inferencia sobre la μ desconocida utiliza esta distribución de $\bar{x}$. La figura 6.1 presenta esta distribución. Los distintos valores de $\bar{x}$ aparecen a lo largo del eje de las abscisas de la figura y la curva normal indica la probabilidad de estos valores.

$$\left.\begin{array}{ccc}
\text{Población} & \xrightarrow{\text{MAS} \quad n=840} & \bar{x} = 272 \\
\mu = ? & \xrightarrow{\text{MAS} \quad n=840} & \bar{x} = 268 \\
\sigma = 60 & \xrightarrow{\text{MAS} \quad n=840} & \bar{x} = 273 \\
 & \vdots & \vdots
\end{array}\right\}$$

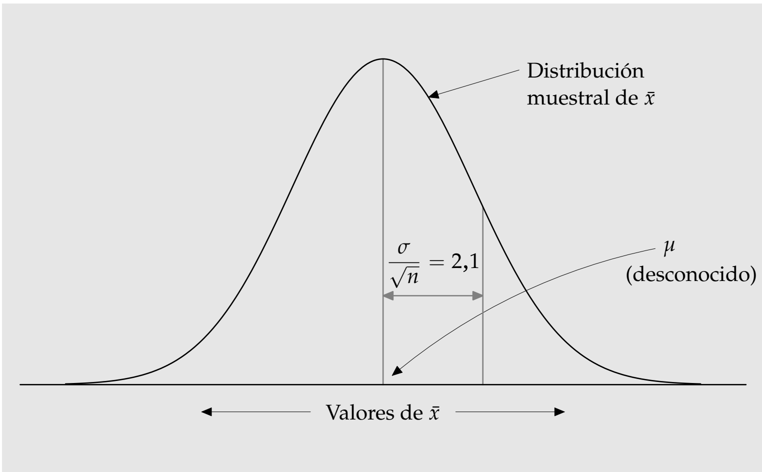


Figura 6.1. Distribución de la puntuación media $\bar{x}$ de la prueba de aritmética en la encuesta NAEP, de una muestra aleatoria simple (MAS) de 840 hombres jóvenes.

6.2.1 Confianza estadística

La figura 6.2 es otra representación de la misma distribución muestral. Esta figura ilustra las siguientes ideas:

- La regla del 68-95-99,7 establece que en un 95% de las muestras $\bar{x}$ se encontrará a menos de dos desviaciones típicas de la media poblacional μ. Por tanto, la media $\bar{x} = 840$ de la prueba NAEP se hallará a menos de 4,2 puntos de μ, en el 95% de todas las muestras.
- Siempre que $\bar{x}$ esté situada a menos de 4,2 puntos de la μ desconocida, la media μ se halla a menos de 4,2 puntos de la $\bar{x}$ observada. Esto ocurrirá en el 95% de todas las muestras.
- Por tanto, en el 95% de las muestras, la μ desconocida está entre $\bar{x} - 4{,}2$ y $\bar{x} + 4{,}2$. La figura 6.3 muestra este hecho de forma gráfica.

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 sobre lo que ocurriría después de muchas repeticiones para expresar nuestra confianza en los resultados de cualquier muestra.

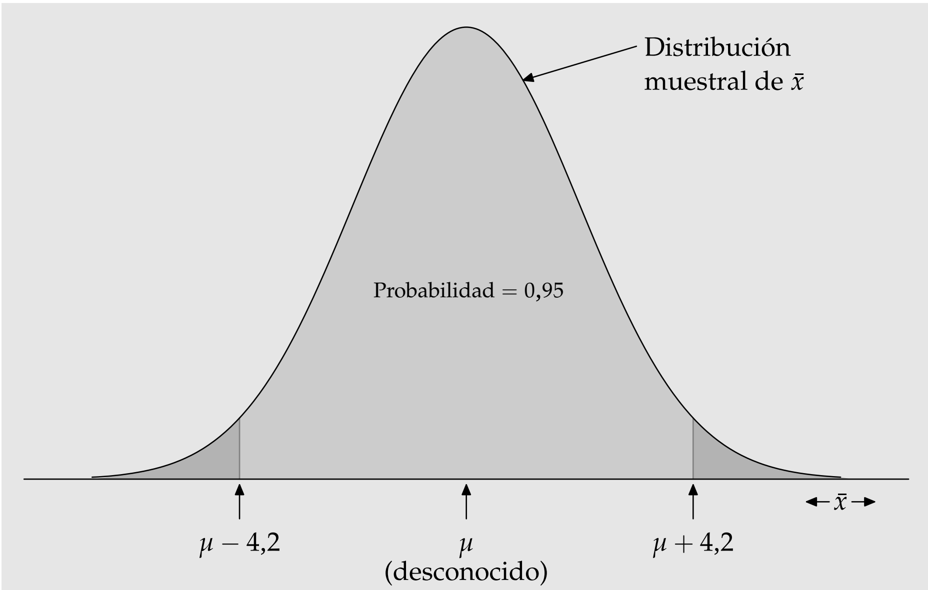


Figura 6.2. En un 95% de las muestras, $\bar{x}$ se encuentra dentro del intervalo $\mu \pm 4{,}2$. Por tanto, μ se encuentra también dentro del intervalo $\bar{x} \pm 4{,}2$ de estas muestras.

$$\begin{array}{lcl}
\text{Población} & \xrightarrow{\text{Muestra aleatoria simple} \quad n=840} & \bar{x} \pm 4{,}2 = 272 \pm 4{,}2 \\
\mu = ? & \xrightarrow{\text{Muestra aleatoria simple} \quad n=840} & \bar{x} \pm 4{,}2 = 268 \pm 4{,}2 \\
\sigma = 60 & \xrightarrow{\text{Muestra aleatoria simple} \quad n=840} & \bar{x} \pm 4{,}2 = 273 \pm 4{,}2 \\
 & \vdots & \vdots
\end{array}\left.\vphantom{\begin{array}{l}a\\a\\a\\a\end{array}}\right\}\ \text{El 95\% de estos intervalos contienen } \mu$$

Figura 6.3. Decir que $\bar{x} \pm 4{,}2$ es un intervalo de confianza para la media poblacional μ significa que en un muestreo repetido, el 95% de estos intervalos contendrán μ.

EJEMPLO 6.2. Confianza del 95%

Nuestra muestra de 840 hombres jóvenes 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

$$\bar{x} - 4{,}2 = 272 - 4{,}2 = 267{,}8 \qquad \text{y} \qquad \bar{x} + 4{,}2 = 272 + 4{,}2 = 276{,}2$$

■

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 del 95% de muestras para las cuales el intervalo $\bar{x} \pm 4{,}2$ contiene μ o, en cambio, es una de las desafortunadas que constituyen 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".

El intervalo 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

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

La estimación ($\bar{x}$ en este caso) es el valor que le suponemos al parámetro desconocido. El **error de estimación** $\pm 4{,}2$ 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.

Error de estimación

INTERVALO DE CONFIANZA

Un **intervalo de confianza de nivel *C*** para un parámetro poblacional tienen dos partes:

- Un intervalo calculado a partir de los datos, en general, tiene la forma

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

- Un **nivel de confianza** C, que proporciona la probabilidad de que en un muestreo repetido, el intervalo contenga el verdadero valor del parámetro.

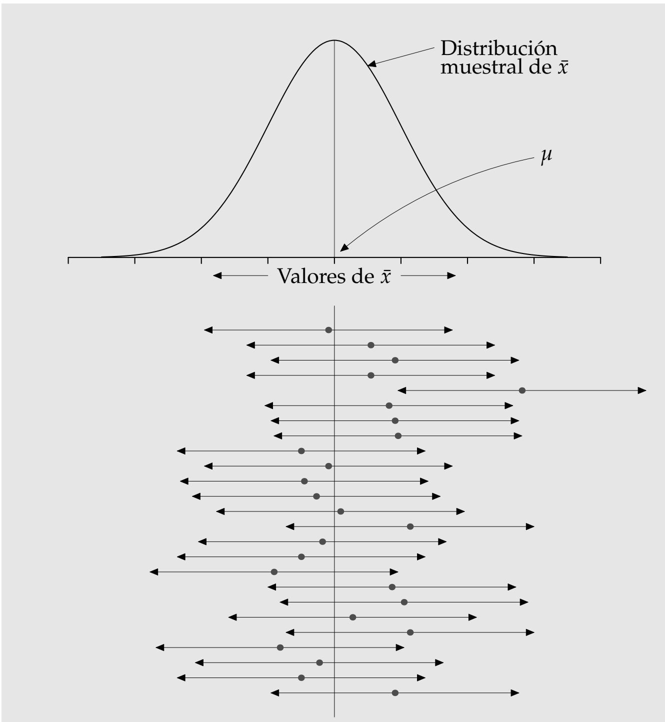


Figura 6.4. Veinticinco muestras de la misma población dieron estos intervalos de confianza del 95%. Después de muchos muestreos, un 95% de las muestras dan intervalos que contienen la media poblacional μ.

Los usuarios pueden elegir el nivel de confianza, en general del 90% o mayor para estar seguros de nuestras conclusiones. Utilizaremos la letra C para indicar el nivel de confianza en tanto por uno. Por ejemplo, un nivel de confianza del 95% corresponde a $C = 0{,}95$.

La figura 6.3 es una manera de representar la idea de un intervalo de confianza del 95%. La figura 6.4 ilustra esta idea de otra forma. Estudia estas figuras con detalle. Si comprendes lo que representan, habrás captado una de las ideas importantes de la estadística. La figura 6.4 presenta el resultado de obtener muchas muestras aleatorias simples de la misma población y calcular un intervalo de confianza para cada una. 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. Vemos el aspecto de esta distribución después de muchas repeticiones. Los intervalos de confianza del 95% de 25 muestras aleatorias simples aparecen debajo. El centro $\bar{x}$ de cada intervalo se ha señalado con un punto, mientras que 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.

APLICA TUS CONOCIMIENTOS

6.1. Encuestas a mujeres. 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 Hawaii. El 47% de las mujeres encuestadas manifestó no tener suficiente tiempo para ellas.

(a) La encuesta daba en sus conclusiones un error de estimación de ± 3 por ciento con una confianza del 95%. Calcula un 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%".

6.2. ¿Qué entendemos por confianza? 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 responde, "el 95% de todos los hombres jóvenes obtienen resultados entre 267,8 y 276,2". ¿Tiene razón el estudiante? Justifica tu respuesta.

6.3. 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 $\sigma = 60$. La media $\bar{x}$ de los 1.000 resultados variará si repites el muestreo.

(a) La distribución de $\bar{x}$ es aproximadamente normal. Su media es $\mu = 280$. ¿Cuál es el valor de la desviación típica?

(b) Dibuja la curva normal que describa cómo varía $\bar{x}$ en muchas muestras de esta población. Señala su media $\mu = 280$ 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 abscisas de tu gráfico como en la figura 6.2.

(d) Siempre que $\bar{x}$ se sitúe en la zona que has señalado, el verdadero valor de la media de la población, $\mu = 280$, 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 y de un valor de $\bar{x}$ que esté situado fuera (utiliza la figura 6.4 como modelo).

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

6.2.2 Intervalos de confianza para la media μ

El razonamiento utilizado para hallar un intervalo de confianza del 95% para la media poblacional μ, se puede aplicar a cualquier nivel de confianza. Partimos de la distribución de la media muestral $\bar{x}$. Si conocemos μ, podemos estandarizar $\bar{x}$. El resultado es el **estadístico z de una muestra:**

Estadístico z de una muestra

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

El estadístico z nos dice si la $\bar{x}$ observada se halla muy lejos de μ, tomando como unidad de medida la desviación típica de $\bar{x}$. Debido a que $\bar{x}$ tiene una distribución normal, z tiene una distribución normal estandarizada $N(0,1)$.

Para hallar un intervalo de confianza del 95%, señala el 95% del área por debajo de la curva. Para un intervalo de confianza de nivel C, marca el área central C. Llama z^* al punto de la distribución normal estandarizada que marca el inicio del área central C del área total 1 por debajo de la curva.

EJEMPLO 6.3. Confianza del 80%

Para hallar un intervalo de confianza del 80% debemos capturar el 80% central de los valores de la distribución normal de $\bar{x}$. Al capturar ese 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 6.5 ilustra la relación entre z^* y las áreas delimitadas por debajo de la curva. ■

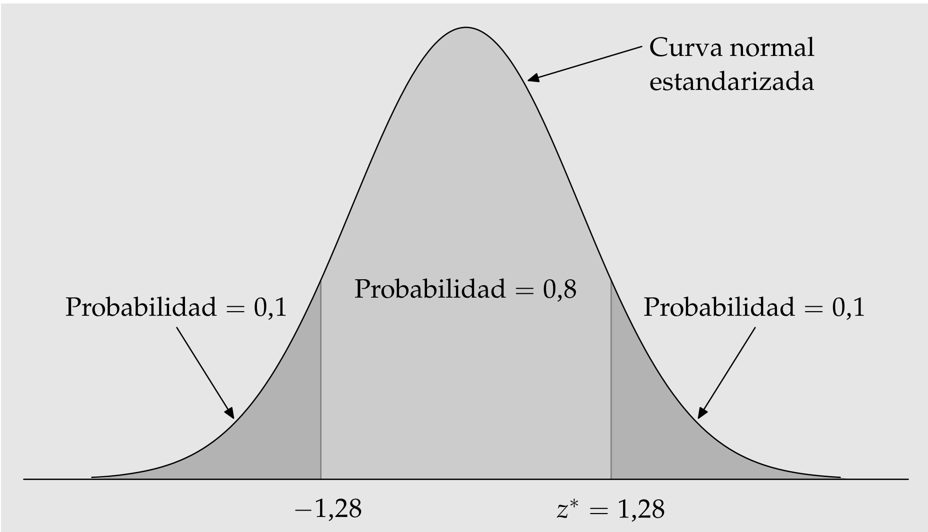


Figura 6.5. La probabilidad central, 0,8 de 1, de la curva normal estandarizada se encuentra entre $-1{,}28$ y 1,28. A la derecha de 1,28 y por debajo de la curva hay un área de 0,1.

La figura 6.6 muestra la relación entre z^* y C: el área correspondiente a C se halla, en la curva normal estandarizada, entre $-z^*$ y z^*. Si nos situamos en la media muestral $\bar{x}$ y nos alejamos z^* desviaciones típicas, obtenemos un intervalo que contiene la media poblacional μ en una proporción C de todas las muestras.

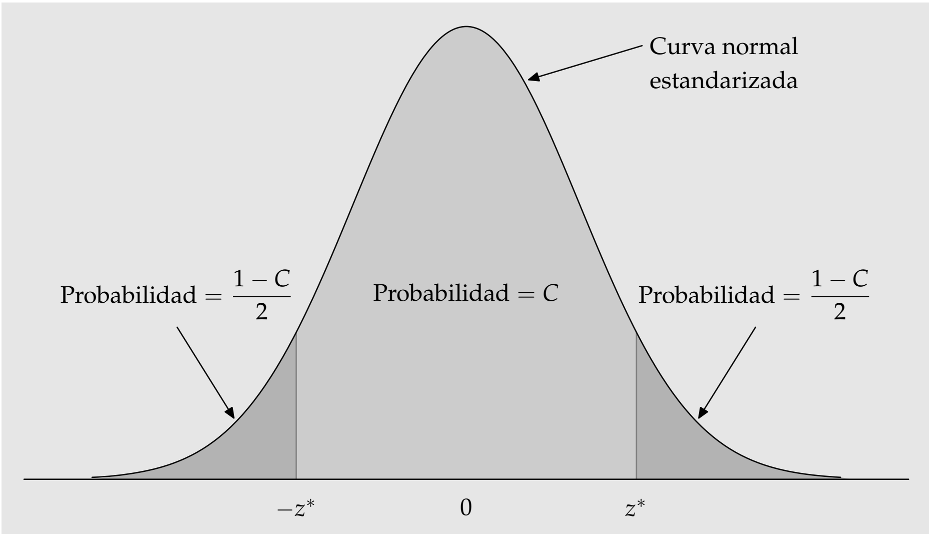


Figura 6.6. El valor crítico z^* es el valor que captura la probabilidad central C por debajo de la curva normal estandarizada entre $-z^*$ y z^*.

Este intervalo es desde $\quad \bar{x} - z^* \frac{\sigma}{\sqrt{n}} \quad$ a $\quad \bar{x} + z^* \frac{\sigma}{\sqrt{n}}, \quad$ es decir,

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

Es un intervalo de confianza para μ del 95%.

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 $z^* = 1{,}960$. Este valor 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.) Los valores de z^* que señalan una determinada área por debajo de la curva normal estandarizada se llaman a menudo **valores críticos** de la distribución.

Valores críticos

INTERVALO DE CONFIANZA PARA LA MEDIA POBLACIONAL

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

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

El valor crítico z^* se ilustra en la figura 6.6 y se puede hallar en la tabla C. El valor de z^* es exacto cuando la distribución poblacional es normal, en los restantes casos es aproximadamente correcto cuando n es grande.

EJEMPLO 6.4. Análisis de medicamentos

Un fabricante de productos farmacéuticos analiza un comprimido de cada uno de los lotes de un medicamento para verificar su concentración de materia activa. El método de análisis químico no es totalmente preciso. Los análisis repetidos de un mismo comprimido dan resultados ligeramente distintos; además, 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

$$0{,}8403 \quad 0{,}8363 \quad 0{,}8447$$

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

$$\begin{aligned} \bar{x} \pm z^* \frac{\sigma}{\sqrt{n}} &= 0{,}8404 \pm 2{,}576 \frac{0{,}0068}{\sqrt{3}} \\ &= 0{,}8404 \pm 0{,}0101 \\ &= 0{,}8303 \text{ a } 0{,}8505 \end{aligned}$$

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 6.4. Repitiendo el cálculo anterior pero para $n = 1$, obtenemos que el intervalo de confianza del 99% basado en un único análisis es

$$\begin{aligned} \bar{x} \pm z^* \frac{\sigma}{\sqrt{1}} &= 0{,}8404 \pm (2{,}576)(0{,}0068) \\ &= 0{,}8404 \pm 0{,}0175 \\ &= 0{,}8229 \text{ a } 0{,}8579 \end{aligned}$$

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 6.7 ilustra la ganancia de precisión cuando se utilizan tres observaciones.

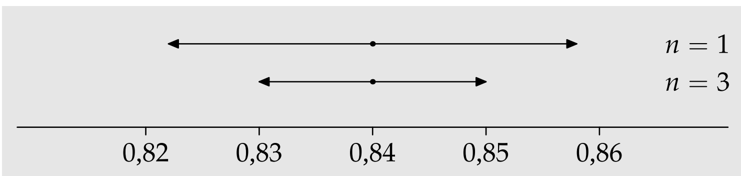


Figura 6.7. Los intervalos de confianza para $n = 1$ y $n = 3$ del ejemplo 6.4. Muestras mayores dan intervalos de confianza más cortos.

APLICA TUS CONOCIMIENTOS

6.4. Encuesta a directores de hotel. 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. La media de tiempo que estos 114 directores habían pasado en su empresa actual era de 11,78 años. Calcula 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.)

6.5. Coeficientes de inteligencia. He aquí los resultados de la prueba IQ de 31 estudiantes de primero de bachillerato:[2]

114	100	104	89	102	91	114	114	103	105	
108	130	120	132	111	128	118	119	86	72	
111	103	74	112	107	103	98	96	112	112	93

(a) Creemos que la distribución de los coeficientes de inteligencia IQ es aproximadamente normal. Dibuja un diagrama de tallos par ver la distribución de los 31 resultados (divide los tallos). Tu diagrama, ¿muestra la presencia de observaciones atípicas, asimetrías claras u otros signos de falta de normalidad?

(b) Trata a las 31 estudiantes como si fueran una muestra aleatoria simple de todas las estudiantes de primero de bachillerato de tu ciudad. Supón que la desviación típica de la distribución de los IQ de esta población es conocida, concretamente $\sigma = 15$. Calcula un intervalo de confianza del 99% para la media del IQ de esta población.

(c) En realidad, los resultados que tenemos corresponden a todas las estudiantes de primero de bachillerato de una determinada escuela de tu ciudad. Explica detalladamente por qué el intervalo de confianza que obtuviste en (b) no es fiable.

6.6. Análisis de sangre. El análisis del nivel de potasio en la sangre no es absolutamente preciso. Además, este nivel en 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 $x = 3{,}2$. Calcula un intervalo de confianza del 90% para la media de su nivel de potasio.

[2]Datos de Darlene Gordon, Purdue University.

(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?

6.2.3 Comportamiento de los intervalos de confianza

El intervalo de confianza $\bar{x} \pm z^* \frac{\sigma}{\sqrt{n}}$ 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 6.6). 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 medio 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 6.5. Cambio del error de estimación

Supón que el fabricante de productos farmacéuticos del ejemplo 6.4 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

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

Al pasar de una confianza del 99% a una del 90%, el error de estimación se ha reducido de $\pm 0{,}0101$ a $\pm 0{,}0065$. La figura 6.8 compara estos dos intervalos.

Aumentar el número de observaciones de 3 a 12 reduce, también, la amplitud del intervalo de confianza del 99% del ejemplo 6.4. 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. ■

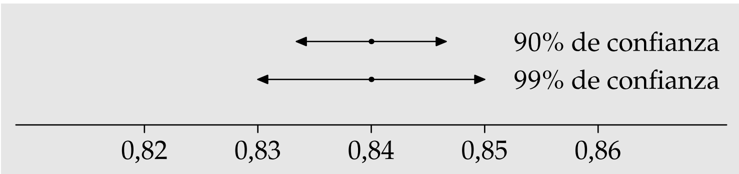


Figura 6.8. Los intervalos de confianza del 90% y del 99% del ejemplo 6.5. Mayor confianza requiere mayor amplitud del intervalo.

APLICA TUS CONOCIMIENTOS

6.7. Nivel de confianza y anchura del intervalo. Los ejemplos 6.4 y 6.5 dan intervalos de confianza para la concentración media 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 6.8 que compare los cuatro intervalos. ¿Cómo afecta el aumento del nivel de confianza a la amplitud del intervalo de confianza?

6.8. Nivel de confianza y error de estimación. La prueba de aritmética de la encuesta NAEP (ejemplo 6.1) 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) Calcula un intervalo de confianza del 95% para la media de los resultados μ de la población de todas las mujeres jóvenes.

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

(c) ¿Cuáles son los errores de estimación con 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?

6.9. Tamaño de muestra y error de estimación. 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) Calcula 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. Calcula un intervalo de confianza del 95% para la media de la población μ en este caso.

(c) A continuación, supón que una muestra de 4.000 mujeres hubiera dado la media muestral $\bar{x} = 275$ y calcula, 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?

6.2.4 Elección del tamaño de la muestra

Un usuario de los métodos estadísticos que sea prudente nunca planifica la obtención de los datos sin planear, al mismo tiempo, la inferencia. Si obtienes suficientes observaciones, entonces 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 $m = z^* \frac{\sigma}{\sqrt{n}}$.

$\bar{x}$ Para obtener el error de estimación deseado m, sustituye el valor de z^* por el que le corresponde según el nivel de confianza escogido y despeja n en la ecuación anterior. Éste el resultado.

TAMAÑO DE LA MUESTRA PARA UN ERROR DE ESTIMACIÓN DESEADO

En un intervalo de confianza para la media poblacional μ, el tamaño de la muestra necesario para un error de estimación m determinado es

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

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 sea mucho mayor que la muestra) no influye sobre el tamaño de la muestra que necesitamos.

EJEMPLO 6.6. ¿Cuántas observaciones?

La gerencia de la empresa exige al laboratorio del ejemplo 6.4 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,

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

Ya 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 para hallar n. Cuando la gerencia conozca el coste de realizar tantos análisis, puede ser que reconsidere su demanda. ■

APLICA TUS CONOCIMIENTOS

6.10. Determinación de la precisión. Para valorar la precisión de una balanza de laboratorio, se pesa repetidamente una pieza con una masa conocida 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 pieza. El resultado medio es de 10,0023 gr. Calcula un intervalo de confianza del 98% para la media de las medidas repetidas de la pieza.

(b) Con una confianza del 98%, ¿cuántas veces se necesita pesar para tener un error de estimación de $\pm 0{,}0001$?

6.11. Con una confianza del 99%, ¿cuál debería ser el tamaño de la muestra de los directores de hotel del ejercicio 6.4 para estimar la media μ con una precisión de ± 1 año?

6.12. Con una confianza del 99%, ¿cuál debería ser el tamaño de la muestra de las estudiantes del ejercicio 6.5 para estimar la media μ con una precisión de ± 5?

6.2.5 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 harían recomendaciones muy detalladas. La fórmula que ya conocemos para la estimación de una media normal, $\bar{x} \pm z^* \frac{\sigma}{\sqrt{n}}$, va acompañada de la siguiente lista de advertencias para el usuario.

- Los datos deben proceder de una muestra aleatoria simple de la población. Estamos completamente a salvo si los datos proceden de una muestra de este tipo. 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 6.4, 6.5 y 6.6, 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 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 ésta 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 $\bar{x} \pm z^* \frac{\sigma}{\sqrt{n}}$ 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 $\bar{x} \pm z^* \frac{s}{\sqrt{n}}$ 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. Ya que muchas de ellas 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, en la práctica, nunca se cumplen totalmente. 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 nos dice la confianza estadística. 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 halla 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 comprendida entre 267,8 y 276,2. Los cálculos de probabilidad en la inferencia estadística describen con qué frecuencia el *método* da una respuesta correcta.

APLICA TUS CONOCIMIENTOS

6.13. Debate sobre una encuesta radiofónica. 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 $\bar{x} = 8.740$ € al año, con una desviación típica de las respuestas $s = 1.125$ €. Para una gran muestra como ésta, s se acerca mucho a la desviación típica desconocida de la población σ. La emisora dice que un intervalo de confianza del 95% para el salario medio μ de los concejales que propondrían todos los ciudadanos va de 8.669 a 8.811 €.

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

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

6.14. Usuarios de Internet. Una encuesta sobre los usuarios de Internet halló que el número de usuarios doblaba al de usuarias. Este resultado fue una sorpresa. Ya que encuestas anteriores establecieron que la relación entre hombres y mujeres era de 9 a 1. En el artículo encontramos la siguiente información:

Se mandaron extensas encuestas a más de 13.000 organizaciones usuarias de Internet; el número de encuestas válidas retornadas fue de 1.468. De acuerdo con el Sr. Quarterman, el error de estimación es, con una confianza del 95%, del 2,8%.[3]

(a) ¿Cuál fue la *proporción de respuestas* de esta encuesta? (La proporción de respuesta es el porcentaje de la muestra que responde adecuadamente.)

(b) ¿Crees que el error de estimación es una buena indicación de la precisión de los resultados de la encuesta? Justifica tu respuesta.

6.15. Oraciones en las escuelas de EE UU. 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 Hawaii ... 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 del 3%. En general, las encuestas de opinión trabajan con confianzas del 95%. 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 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 que deban añadirse al error de estimación del $\pm 3\%$. Presta especial atención a la descripción del método de muestreo que da el artículo.

[3] P. H. Lewis, "Technology" column, *New York Times*, de 29 de mayo de 1995.

RESUMEN DE LA SECCIÓN 6.2

Un **intervalo de confianza** utiliza una muestra de datos para estimar un parámetro desconocido con 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, el **valor crítico 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.

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 6.2

6.16. Asistencia social. Un artículo sobre una encuesta señala que el "28% de los 1.548 adultos entrevistados creían que las personas que eran capaces de trabajar no deberían recibir asistencia social". El artículo seguía diciendo: "el error de estimación para una muestra de tamaño 1.548 es de más menos el tres por ciento". A menudo, las encuestas de opinión anuncian errores de estimación para confianzas del 95%. A partir de esta consideración, explica a alguien que no sepa estadística lo que significa "un error de estimación de más menos un tres por ciento".

6.17. Sistemas informáticos en hoteles. ¿Están satisfechos los directores de hotel con los sistemas informáticos que utilizan en sus establecimientos? Se envió una encuesta a 560 directores de hoteles de entre 200 y 500 habitaciones en Chicago y Detroit.[4] Fueron devueltas 135. 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 con la facilidad de uso de los sistemas informáticos, la media de los resultados fue $\bar{x} = 5{,}396$. Calcula un intervalo de confianza del 95% para la media poblacional. (Parte del supuesto de 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 $\sigma = 1{,}75$, calcula un intervalo de confianza del 99% para la media poblacional.

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

(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?

6.18. ¿Por qué la gente prefiere las farmacias? 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 la atención en los tres tipos de establecimientos, utilizando un extenso cuestionario que preguntaba cosas 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 en la guía telefónica de Indianápolis. He aquí las medias y las desviaciones típicas de las puntuaciones otorgadas por los sujetos de la muestra.[5]

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, en general, 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?

(b) Calcula 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?

6.19. Curación de heridas en la piel. Unos investigadores que estudiaban la cicatrización de las heridas de la piel midieron la rapidez con la que se cerraba un

[5] Datos proporcionados por Mugdha Gore y Joseph Thomas, Purdue University School of Pharmacy.

corte hecho con una hoja de afeitar en la piel de un tritón anestesiado. He aquí los resultados de 18 tritones, expresados en micras (millonésima parte de un metro) por hora:[6]

29	27	34	40	22	28	14	35	26
35	12	30	23	18	11	22	23	33

(a) Dibuja un diagrama de tallos con estos datos (divide los tallos). Es difícil valorar la normalidad a partir de 18 observaciones; sin embargo, busca si hay observaciones atípicas o asimetrías extremas. ¿Qué hallaste?

(b) En general, los científicos dan por supuesto que los animales de la muestra constituyen una muestra aleatoria simple de su especie o tipo genético. Considera que estos tritones son una muestra aleatoria simple y supón, además, que la desviación típica poblacional de la velocidad de cierre de las heridas de esta especie es de 8 micras por hora. Calcula un intervalo de confianza para la media de la velocidad de cierre de esta especie.

(c) Una amiga que casi no sabe nada de estadística utiliza la fórmula $\bar{x} \pm z^* \frac{\sigma}{\sqrt{n}}$ que ha sacado de un manual de biología para calcular un intervalo de confianza del 95% para la media. Su intervalo de confianza, ¿es más ancho o más estrecho que el nuestro? Explica a tu amiga por qué una mayor confianza cambia la anchura del intervalo.

6.20. Cigüeñales. 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) Calcula un intervalo de confianza del 95% para la media del proceso en el momento en que se produjeron estos cigüeñales.

[6]D. D. S. Iglesia, E. J. Cragoe, Jr. y J. W. Vanable, "Electric field strength and epithelization in the newt (*Notophthalmus viridescens*)", *Journal of Experimental Zoology*, 274, 1996, págs. 56-62.

6.21. ¿Cuál debería ser el tamaño de la muestra que te permitiría estimar la velocidad media de cicatrización de las heridas en la piel de los tritones (consulta el ejercicio 6.19) con un error de estimación de 1 micra por hora y una confianza del 90%?

6.22. Encuesta de un periódico. 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, exceptuando Alaska y Hawai. La encuesta daba un error de estimación de $\pm 3\%$ para una confianza del 95% con los resultados de las mujeres. El error de estimación de los resultados de los hombres era de $\pm 4\%$. ¿Por qué el error de estimación de los hombres es mayor que el error de estimación de las mujeres?

6.23. Encuesta preelectoral. En 1976 las elecciones presidenciales de EE UU, 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 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.

6.24. Cómo se hizo la encuesta. El *New York Times* incluye un recuadro titulado "Cómo se hizo la encuesta" en artículos basados en sus propias encuestas de opinión. A continuación se presentan fragmentos de unos de estos recuadros (marzo de 1995). En el recuadro se habla de un error de estimación de más menos el tres por cien, con una confianza del 95%.

La última encuesta del New York Times/*CBS se basa en entrevistas telefónicas llevadas a cabo entre el 9 y el 12 de marzo a 1.156 adultos de todos los EE UU, excepto*

Alaska y Hawai. (A continuación se describe el método de obtención al azar de números de teléfono utilizado en la encuesta.)

Además del error de muestreo, las dificultades prácticas inherentes a la realización de un encuesta de opinión pública pueden introducir otras fuentes de error en la encuesta. Por ejemplo, la manera de formular las preguntas o el orden de las mismas pueden conducir a resultados distintos.

(a) El párrafo anterior menciona diferentes fuentes de error en los resultados de la encuesta. Haz un listado de estas fuentes de error.

(b) ¿Cuál de las fuentes de error que mencionaste en (a) queda incluida dentro del error de estimación anunciado?

6.3 Pruebas de significación

Los intervalos de confianza son uno de los dos procedimientos de inferencia estadística más ampliamente utilizados. Úsalos cuando tu objetivo sea estimar un parámetro poblacional. El segundo procedimiento de inferencia más ampliamente utilizado, llamado *pruebas de significación*, tiene otro objetivo: valorar la evidencia proporcionada por los datos a favor de alguna hipótesis sobre la población.

EJEMPLO 6.7. Soy un gran encestador de tiros libres

Puedo encestar el 80% de los tiros libres que lanzo. Para comprobar mi afirmación, me pides que lance 20 tiros libres. Solamente encesto 8 de 20. "¡Aja!", exclamas, "alguien que encesta el 80% de los tiros libres, casi nunca encestaría sólo 8 de 20. Por tanto, no te creo".

Tu razonamiento se basa en preguntar lo que ocurriría si mi afirmación fuera correcta y repitiéramos el lanzamiento de 20 tiros libres muchas veces —casi nunca encestaría sólo 8—. Este resultado es tan poco probable que proporciona una fuerte evidencia de que mi afirmación no es cierta.

Puedes determinar la fuerza de la evidencia en contra de mi afirmación calculando la probabilidad de que sólo enceste 8 tiros de 20 si realmente encesto el 80% después de muchas repeticiones. Esta probabilidad es de 0,0001. Solamente encestaría 8 de 20 una vez de cada 10.000 intentos de 20 lanzamientos si mi afirmación del 80% fuera cierta. Esta probabilidad tan pequeña te convence de que mi afirmación es falsa. ■

Las pruebas de significación utilizan una terminología muy elaborada, de todas formas las idea básica es sencilla: si suponemos que una determinada afirmación es cierta y bajo esta suposición observamos que un determinado suceso ocurre muy raramente, esto indica que la afirmación no es cierta.

6.3.1 Razonamientos de las pruebas de significación

Los razonamientos utilizados con las pruebas de significación, al igual que los utilizados con los intervalos de confianza, se basan en preguntar lo que ocurriría si repitiéramos el muestreo o el experimento muchas veces. Otra vez empezaremos utilizando un procedimiento poco realista con el objetivo de dar énfasis a los razonamientos. He aquí uno de los ejemplos que exploraremos.

EJEMPLO 6.8. Refrescos light

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:

$$2{,}0 \quad 0{,}4 \quad 0{,}7 \quad 2{,}0 \quad -0{,}4 \quad 2{,}2 \quad -1{,}3 \quad 1{,}2 \quad 1{,}1 \quad 2{,}3$$

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?* ■

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

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

Suponemos, de forma poco realista, que conocemos que la desviación típica de la población de catadores es $\sigma = 1$.

Los razonamientos son como los del ejemplo 6.7. Afirmamos algo y nos preguntamos si los datos dan evidencia *en contra* de lo que afirmamos. Buscamos si hay evidencia de que hay una pérdida de dulzura; por tanto, la afirmación que contrastamos es que *no* hay una pérdida de dulzura. En tal caso, la pérdida media de dulzura detectada por la población de todos los catadores sería $\mu = 0$.

- Si la afirmación de que $\mu = 0$ es cierta, la distribución de $\bar{x}$ de una muestra de 10 catadores es normal con media $\mu = 0$ y desviación típica

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

  La figura 6.9 muestra esta distribución muestral. Podemos valorar si cualquier valor observado de $\bar{x}$ es sorprendente situándolo en esta distribución.
- Supón que nuestros 10 catadores dieron una media de pérdida de dulzura $\bar{x} = 0{,}3$. Queda claro a partir de la figura 6.9 que un valor de $\bar{x}$ como ésta puede ocurrir sólo por azar cuando la media de la población es $\mu = 0$. Que 10 catadores dieran una $\bar{x} = 0{,}3$ no constituye ninguna evidencia de pérdida de dulzura.
- En realidad, el resultado de nuestra prueba dio una $\bar{x} = 1{,}02$. Este valor de $\bar{x}$ queda muy alejado de $\mu = 0$ en la curva normal de la figura 6.9, 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.

6.3.2 Terminología de las pruebas de significación

Una prueba estadística empieza con un cuidadoso planteamiento de las afirmaciones que estamos interesados en comparar. Estas afirmaciones hacen referencia

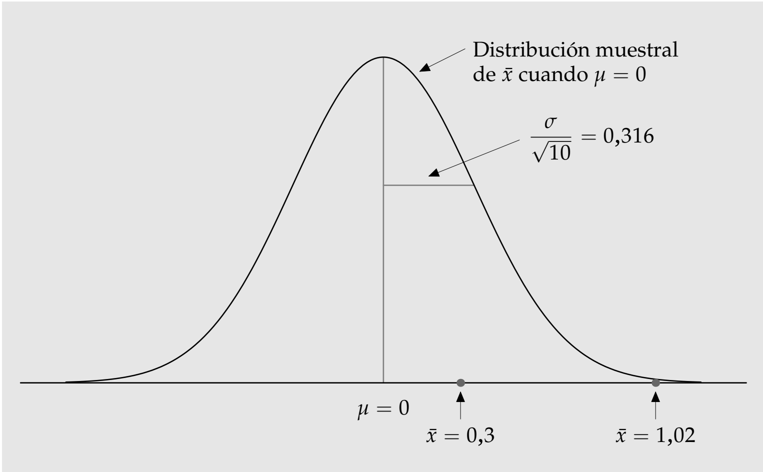


Figura 6.9. Si un refresco no pierde dulzura durante su almacenamiento, la puntuación media $\bar{x}$ de 10 catadores tendrá esta distribución muestral. El resultado de un refresco fue $\bar{x} = 0{,}3$. Esto podría ocurrir muy fácilmente por azar. Otro refresco tenía $\bar{x} = 1{,}02$. Este valor se encuentra muy alejado de $\mu = 0$ en la distribución normal y constituye, por tanto, una buena evidencia a favor de que dicho refresco ha perdido dulzura.

a una población; por tanto, las expresamos en términos de parámetros poblacionales. En el ejemplo 6.8, el parámetro es la media poblacional μ, la pérdida media de dulzura que detectaría una muestra muy grande de catadores. Debido a que los razonamientos que hemos señalado buscan evidencia *en contra* de la afirmación, empezamos enunciando la frase en contra de la que buscamos evidencia, como por ejemplo "no hay pérdida de dulzura". Esta afirmación es nuestra *hipótesis nula*.

HIPÓTESIS NULA H_0

La afirmación que se contrasta en una prueba estadística se llama **hipótesis nula.** Las pruebas de significación se diseñan para 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 afirmación en relación con la población sobre la cual queremos hallar evidencia *a favor* es la **hipótesis alternativa**, designada como H_a. En el ejemplo 6.8, buscamos evidencia de que hay una pérdida de dulzura. La hipótesis nula dice "no hay pérdida" como media en una gran población de catadores. La hipótesis alternativa dice "hay pérdida". Por tanto, las hipótesis son

Hipótesis alternativa

$$H_0 : \mu = 0$$

$$H_a : \mu > 0$$

Las hipótesis nula y alternativa son planteamientos precisos sobre las afirmaciones que contrastamos. Si obtuviéramos un resultado que fuera poco probable si H_0 fuera cierta en la dirección propuesta por la H_a, tendríamos evidencia en contra de H_0 y a favor de H_a.

VALOR *P*

La probabilidad, calculada suponiendo que H_0 es cierta, de que el resultado tome un valor al menos tan extremo como el observado se llama **valor *P*** de la prueba de significación. Cuanto menor sea el valor P, más fuerte es la evidencia que proporcionan los datos en contra de H_0.

EJEMPLO 6.9. ¿Qué es el valor P?

La figura 6.10 muestra el valor P cuando 10 catadores dan una pérdida media de dulzura $\bar{x} = 0{,}3$. Es la probabilidad de que obtengamos una media muestral al menos tan grande como 0,3 en el caso de que $\mu = 0$ fuera *cierto*. Esta probabilidad es $P = 0{,}1711$. Es decir, obtendríamos una pérdida de dulzura como ésta o mayor aproximadamente el 17% de las veces, sólo por azar al escoger a 10 catadores, incluso si la media de toda la población de catadores no hallara pérdida de dulzura. Un resultado que ocurre con esta frecuencia cuando H_0 es cierta, no es una buena evidencia en contra de H_0.

En realidad, la media de los 10 catadores fue $\bar{x} = 1{,}02$. El valor P es la probabilidad de obtener una $\bar{x}$ al menos tan grande si en realidad $\mu = 0$. Esta probabilidad es $P = 0{,}0006$. Raramente obtendríamos una muestra con una pérdida de dulzura tan grande si la H_0 fuera cierta. El valor P tan pequeño proporciona una fuerte evidencia en contra de H_0 y en cambio una fuerte evidencia a favor de la alternativa $H_a : \mu > 0$. ■

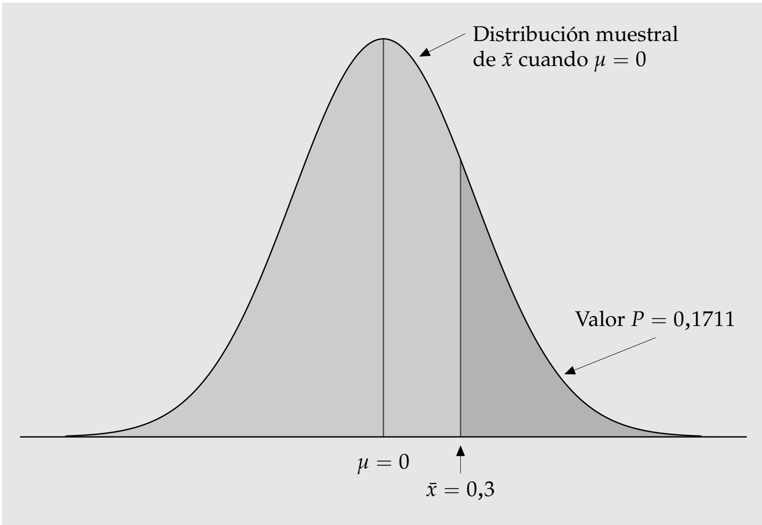


Figura 6.10. El valor P del resultado $\bar{x} = 0{,}3$ en la prueba de cata de refrescos. El valor P es la probabilidad (cuando H_0 es cierta) de que $\bar{x}$ tome un valor al menos tan grande como el realmente observado.

Valores P pequeños proporcionan evidencia en contra de H_0, ya que nos dicen que es poco probable que el resultado obtenido ocurra sólo por azar. Valores P grandes no proporcionan evidencia en contra de H_0.

Las recetas sobre las pruebas de significación no dejan translucir los razonamientos que hay detrás. En realidad los programas estadísticos, a menudo sólo proporcionan el valor P. Observa otra vez las $\bar{x}$ sobre la pérdida de dulzura de los dos refrescos de la figura 6.9. Podemos ver que uno de los resultados no es sorprendente si la verdadera media poblacional es 0. En cambio, el otro resultado sí lo es. Una prueba de significación dice lo mismo pero de forma más detallada.

APLICA TUS CONOCIMIENTOS

6.25. Actitud de los estudiantes. 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 y 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) Si la hipótesis nula es cierta, ¿cuál es la distribución de la media de los resultados $\bar{x}$ de una muestra de 25 estudiantes mayores de 30 años? 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 abscisas 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 abscisas de tu gráfico. En realidad, el resultado fue $\bar{x} = 125{,}7$. Señala este punto en tu dibujo y utilizandolo, 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 $\bar{x} = 118{,}6$.

6.26. Gastos dedicados a la vivienda. La Oficina del Censo de EE UU informa que los hogares de este país dedicaron una media del 31% de todos sus gastos a la vivienda. Una asociación de constructores de la ciudad de Cleveland cree que esta media es menor en su zona y entrevistan a 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 poblacional 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 abscisas lo que sabes sobre las posiciones 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 abscisas de tu dibujo. Luego, supón que el resultado del estudio es $\bar{x} = 27{,}6\%$. Señala este punto en tu dibujo. Refiriéndote a él, explica con un lenguaje sencillo por qué uno de los resultados constituye una buena evidencia a favor de que la media del gasto en la 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 $\bar{x} = 30{,}2\%$. (Fíjate en que estamos buscando evidencia a favor de que el gasto es menor que el que supone la hipótesis nula.)

6.3.3 Más detalles: planteamiento de las hipótesis

En una prueba de significación primero planteamos las hipótesis. La hipótesis nula es una afirmación *en contra* de la cual intentaremos encontrar evidencia. La hipótesis alternativa H_a es una afirmación sobre la población *a favor* de la cual intentamos encontrar evidencia. En el ejemplo 6.8 estuvimos buscando evidencia a favor de una pérdida de dulzor. La hipótesis nula establece que como media "no hay pérdida" de dulzor 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. Veamos otro ejemplo.

EJEMPLO 6.10. Estudio de satisfacción en el trabajo

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. Este experimento constituye otro ejemplo de diseño por pares. La variable respuesta es la diferencia entre las puntuaciones después de estar trabajando en la línea de montaje al ritmo de trabajo marcado por las operarias y después de estar trabajando al ritmo que marca una máquina.[7]

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. La hipótesis nula dice que no hay diferencias entre las dos condiciones de trabajo, es decir,

$$H_0 : \mu = 0$$

Los autores del estudio estaban interesados en 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**,

Alternativa de dos colas

$$H_a : \mu \neq 0$$

■

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 6.10 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,

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

sólo estamos interesados en detectar variaciones positivas de μ. La hipótesis alternativa debe expresar nuestras sospechas o nuestras esperanzas sobre los datos. Sería 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 6.10 estuvieran más satisfechas cuando ellas mismas marcaban el ritmo de trabajo no tiene que influir en la elección de H_a. Si no has tomado por adelantado una decisión firme sobre la dirección del efecto, utiliza una alternativa de dos colas.

APLICA TUS CONOCIMIENTOS

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.

6.27. Se supone que el diámetro de un eje de un pequeño motor 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.

6.28. Los datos de la Oficina del Censo 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 52.500 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 de la población general.

6.29. 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 del curso, así que el catedrático compara la media de esos estudiantes con la media general del grupo.

6.30. El año pasado el servicio técnico de tu empresa dedicó una media 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 una distinta media del tiempo dedicado a resolver problemas por teléfono?

6.3.4 Más detalles: valores P y significación estadística

Una prueba de significación utiliza los datos en forma de **estadístico de contraste**. Éste 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 $\bar{x}$.

Estadístico de contraste

Una prueba de significación valora la evidencia en contra de la hipótesis nula en términos de probabilidad, el valor P. 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. El valor P describe la fuerza de la evidencia, ya que es 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.

EJEMPLO 6.11. Cálculo del valor P de una cola

En el ejemplo 6.8 las observaciones son una muestra aleatoria simple de tamaño $n = 10$ de una población normal con $\sigma = 1$. La pérdida media de dulzor observada en un refresco fue $\bar{x} = 0{,}3$. El valor P para contrastar es

$$H_0 : \mu = 0$$

$$H_a : \mu > 0$$

por tanto,

$$P(\bar{x} \geq 0{,}3)$$

calculado suponiendo que H_0 es cierta. Cuando lo es, $\bar{x}$ 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 por debajo de la curva. La figura 6.11 es el gráfico de este ejemplo. Luego estandariza $\bar{x}$ para tener la distribución normal estandarizada Z y utiliza la tabla A,

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

Este valor P es el que aparece en la figura 6.10. ■

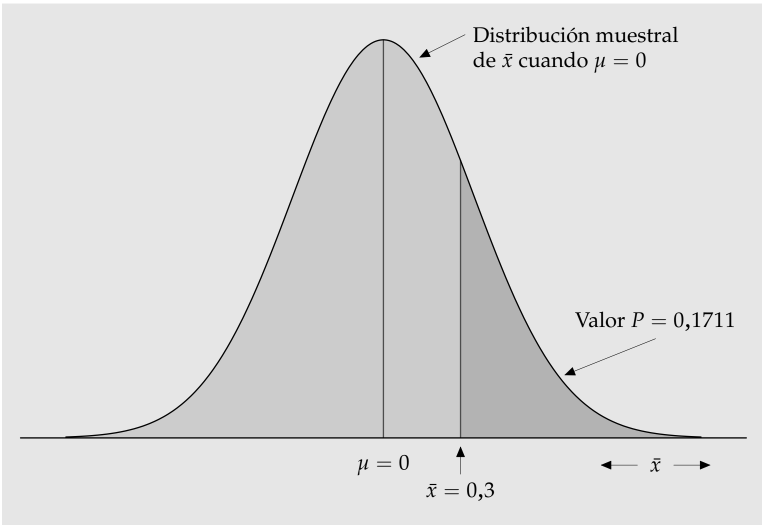


Figura 6.11. El valor P de la prueba de una cola del ejemplo 6.11.

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 P decisivo 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

Nivel de significación

valor de $\bar{x}$ no ocurra por azar más del 5% de las veces (en 1 de cada 20 muestras) cuando H_0 sea 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 ocurre por azar en un 1% de las ocasiones (en 1 de cada 100 muestras) si H_0 es 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". 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 $\alpha = 0{,}01$.

APLICA TUS CONOCIMIENTOS

6.31. Vuelve al ejercicio 6.25.

(a) 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 6.25.

(b) ¿Cuál de los dos valores observados de $\bar{x}$ es estadísticamente significativo al nivel $\alpha = 0{,}05$? ¿Y al nivel $\alpha = 0{,}01$?

6.32. Vuelve al ejercicio 6.26.

(a) 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 6.26.

(b) El valor $\bar{x} = 27{,}6$, ¿es estadísticamente significativo al nivel $\alpha = 0{,}05$? ¿Y al nivel $\alpha = 0{,}01$?

6.33. Ganancias de los estudiantes. La oficina de ayuda financiera de una universidad pregunta a una muestra de estudiantes sobre sus empleos y ganancias. El informe dice que "con relación a las ganancias de los estudiantes durante el curso académico, se halló una diferencia significativa ($P = 0{,}038$) entre sexos; como media, 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 con relación a la influencia del sexo y de la raza sobre la media de las ganancias.[8]

6.3.5 Pruebas de significación para una media poblacional

Al llevar a cabo una prueba de significación, tienes que seguir tres pasos:

1. Plantear las hipótesis.
2. Calcular el estadístico de contraste.
3. Hallar el valor P.

Una vez hayas planteado 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 determinado 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*

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

[8] De un estudio de M. R. Schlatter *et al*., Division of Financial Aid, Purdue University.

Estadístico z de una muestra

Este ***estadístico z de una muestra*** 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 6.11 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 sólo plantea que μ es 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.

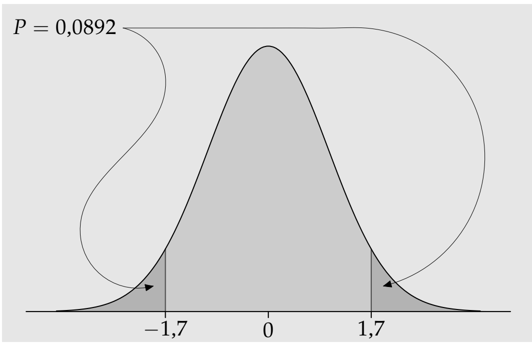


Figura 6.12. El valor P de la prueba de significación de dos colas del ejemplo 6.12.

EJEMPLO 6.12. Cálculo del valor P de dos colas

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 6.12 muestra esta probabilidad como 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.

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

Si el valor observado fuera $z = -1{,}7$, haríamos exactamente los mismos cálculos. Lo que importa es el valor absoluto $|z|$, no si z es positivo o negativo. ■

PRUEBA *z* PARA UNA MEDIA POBLACIONAL

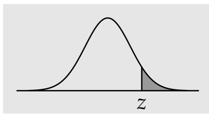
Para contrastar la hipótesis $H_0 : \mu = \mu_0$ a partir de una muestra aleatoria simple de tamaño n de una población con media desconocida y desviación típica σ conocida, calcula el **estadístico de contraste *z***

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

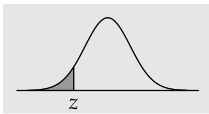
En términos de una variable Z que tiene una distribución normal estandarizada.

Para contrastar H_0 en contra de las siguientes alternativas, los valores P son los siguientes:

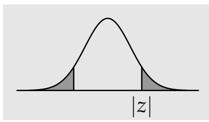
$H_a : \mu > \mu_0$ es $P(Z \geq z)$



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



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



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

EJEMPLO 6.13. Presión de la sangre de los ejecutivos

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 consiguiente, las hipótesis sobre la media desconocida μ de la población de ejecutivos son

$$H_0 : \mu = 128$$

$$H_a : \mu \neq 128$$

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

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

Paso 3: Valor P. Sigue siendo conveniente que hagas un dibujo que te ayude a hallar el valor P, aunque basta con que señales el valor z observado en una curva normal estandarizada. La figura 6.13 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 está situada al menos tan lejos de 128 como la media de la muestra de ejecutivos. 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. ■

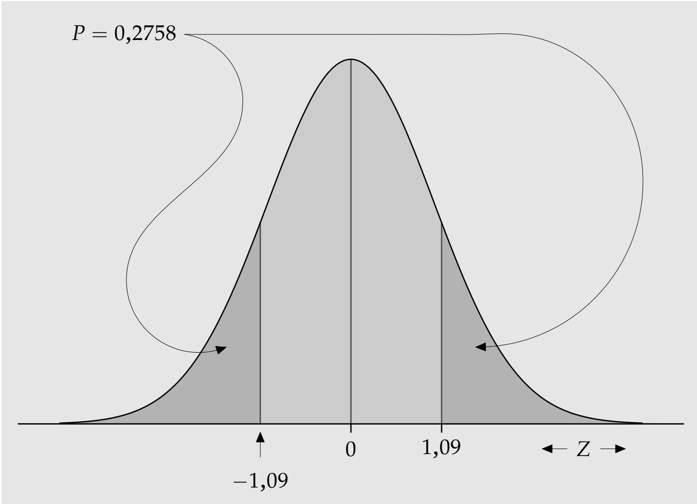


Figura 6.13. El valor P de la prueba de significación de dos colas del ejemplo 6.13

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 ellos pasan cada año una revisión médica, no es difícil tomar una muestra aleatoria simple que represente al colectivo.

Los datos del ejemplo 6.13 *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 es fuerte, podemos estar seguros de que hay que rechazar H_0 en favor de H_a. No hallar evidencia en contra de H_0 sólo significa que los datos son compatibles con H_0, no significa que sea cierto que tengamos una clara evidencia a favor de que H_0 es verdadera.

EJEMPLO 6.14. ¿Eres capaz de determinar el saldo de tu cuenta corriente?

En un debate sobre el nivel de formación de la mano de obra de EE UU, alguien dice, "como media, los jóvenes de hoy no son capaces ni de calcular el saldo de su cuenta corriente". La encuesta NAEP afirma que una persona que obtenga una puntuación de al menos 275 en la prueba de aritmética (repasa el ejemplo 6.1) 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 6.1, 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

$$z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} = \frac{272 - 275}{60/\sqrt{840}} = -1{,}45$$

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 6.14 ilustra el valor P. Utilizando la tabla A, encontramos que

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

Conclusión: Una puntuación media 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. Esta prueba no es significativa a un nivel $\alpha = 0{,}05$. ■

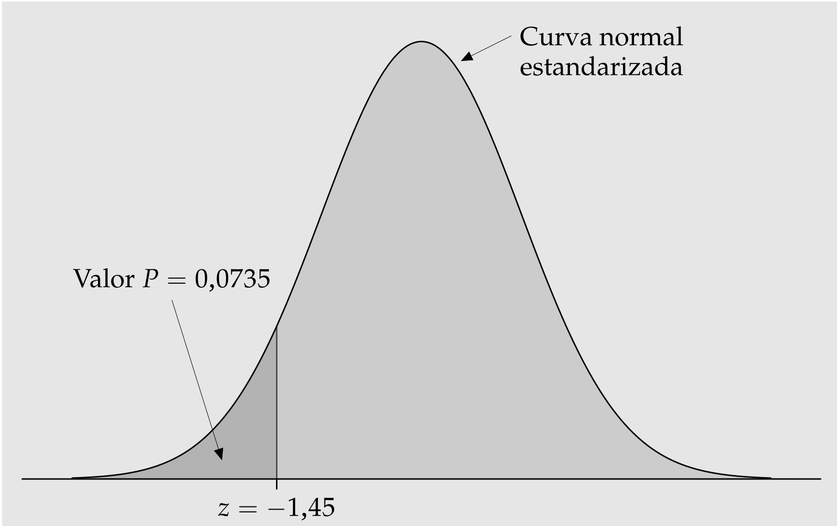


Figura 6.14. El valor P de la prueba de significación de una cola correspondiente al ejemplo 6.14

APLICA TUS CONOCIMIENTOS

6.34. Ventas de café. 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 $\sigma = 33$ 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 la media de ventas es ahora mayor? Las hipótesis son

$$H_0 : \mu = 354$$

$$H_a : \mu > 354$$

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

(a) Halla la media muestral $\bar{x}$ y el valor del estadístico de contraste z.

(b) Calcula el valor P. Dibuja la curva normal y sombrea el área que representa el valor P.

(c) 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?

6.35. Cigüeñales. 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) Halla el valor P de la prueba. ¿Estás convencido de que la media del proceso no es de 224 mm?

6.36. Llenando botellas de cola. Se supone que las botellas de una famosa cola 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 cola es menor de 300 ml?

(a) Plantea las hipótesis que contrastarás.

(b) Calcula el estadístico de contraste.

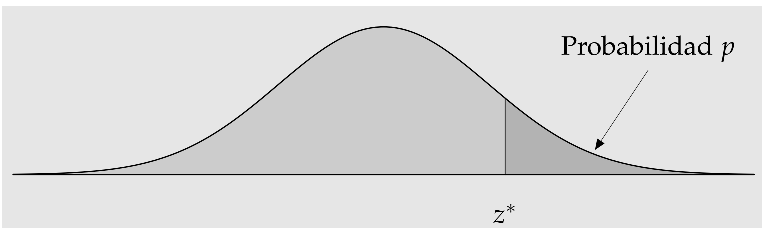
(c) Halla el valor P y expresa tus conclusiones.

6.3.6 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. Sin embargo, primero vamos a describir de forma completa los valores críticos.

VALORES CRÍTICOS

El valor z^* que tiene una probabilidad p a su derecha por debajo de la curva normal estandarizada se llama **valor crítico superior** de la distribución normal estandarizada.



La fila de valores z^* de la tabla C proporciona los valores críticos correspondientes a las probabilidades p de la fila superior de la tabla.

EJEMPLO 6.15. ¿Es significativo?

En el ejemplo 6.14 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 obtenido $z = -1{,}45$ con el valor crítico del 5%, $z^* = 1{,}645$, de la tabla C. Debido a que este valor $z = -1{,}45$ *no* se encuentra más lejos de 0 que $-1{,}645$, la prueba *no* es significativa a un nivel $\alpha = 0{,}05$.

La figura 6.15 muestra cómo $-1{,}645$ separa los valores z que son significativos a un nivel $\alpha = 0{,}05$ de los que no lo son. ■

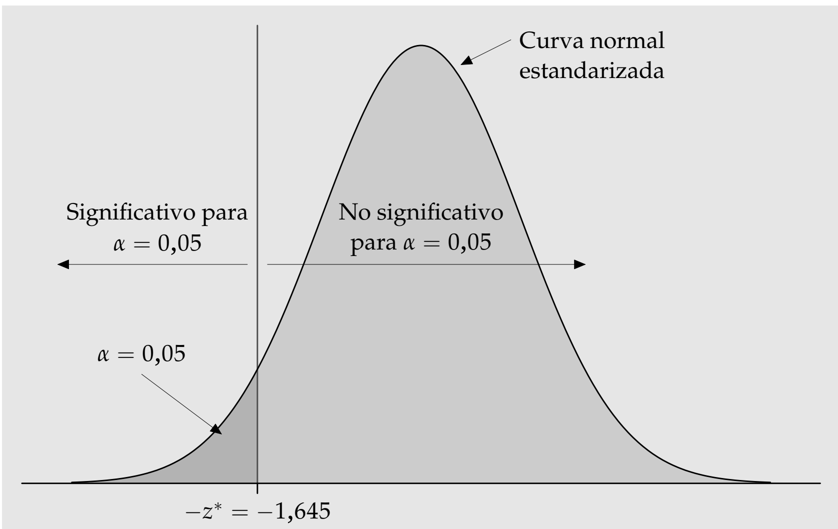


Figura 6.15. Decisión de si el estadístico z es significativo a un nivel de significación $\alpha = 0{,}05$ en la prueba de una cola del ejemplo 6.15.

PRUEBAS CON UN NIVEL DE SIGNIFICACIÓN PREDETERMINADO

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 μ desconocida y desviación típica conocida σ, calcula el estadístico z de contraste

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

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

$$H_a : \mu > \mu_0 \qquad \text{si} \qquad z \geq z^*$$

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

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

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

donde z^* es el valor crítico superior $\frac{\alpha}{2}$ obtenido en la tabla C.

EJEMPLO 6.16. La concentración, ¿es correcta?

Al laboratorio de análisis del ejemplo 6.4 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 $\frac{\alpha}{2} = 0{,}005$ de la tabla C. Este valor crítico es $z^* = 2{,}576$. La figura 6.16 muestra cómo los valores críticos separan los valores z que son estadísticamente significativos de los que no lo son. 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. ■

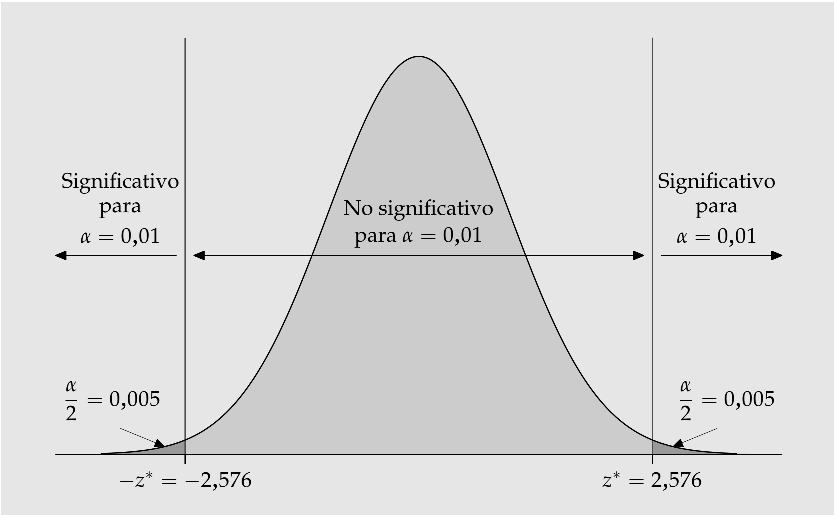


Figura 6.16. Decisión de si el estadístico z es significativo a un nivel de significación $\alpha = 0{,}01$ en la prueba de dos colas del ejemplo 6.16.

El resultado obtenido en el ejemplo 6.16 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

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

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 y conocerlo nos permite valorar la significación a cualquier nivel.

Además de valorar la significación para valores α predeterminados, las tablas de valores críticos, como la tabla C, nos permiten estimar los valores P sin realizar cálculos de probabilidad. En el ejemplo 6.16 se compara el valor observado $z = -4{,}99$ con todos los valores críticos normales de la última fila de la tabla C. 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 6.15, $z = -1{,}45$ 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.

APLICA TUS CONOCIMIENTOS

6.37. Generador de números aleatorios. 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 poblacional 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% ($\alpha = 0{,}01$)?

(d) ¿Entre qué dos valores críticos normales de la última fila de la tabla C se encuentra z? ¿Entre qué dos valores se halla el valor P?

6.38. Nicotina en cigarrillos. 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%?

(c) Comparando z con los valores críticos de la última fila de la tabla C, ¿entre qué dos valores se halla el valor P?

6.3.7 Pruebas derivadas de los intervalos de confianza

Los cálculos del ejemplo 6.16 para una prueba de significación del 1% son muy similares a los del ejemplo 6.4 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 $H_0 : \mu = \mu_0$ cuando el valor μ_0 se halla fuera del intervalo de confianza de nivel $1 - \alpha$ para μ.

EJEMPLO 6.17. Pruebas de significación a partir de un intervalo

El intervalo de confianza del 99% para μ del ejemplo 6.4 es

$$\begin{aligned} \bar{x} \pm z^* \frac{\sigma}{\sqrt{n}} &= 0{,}8404 \pm 0{,}0101 \\ &= 0{,}8303 \text{ a } 0{,}8505 \end{aligned}$$

El valor hipotético $\mu_0 = 0{,}86$ del ejemplo 6.16 se halla fuera de este intervalo de confianza, por tanto 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 halla dentro del intervalo de confianza del 99% para μ. La figura 6.17 ilustra estos dos casos. ■

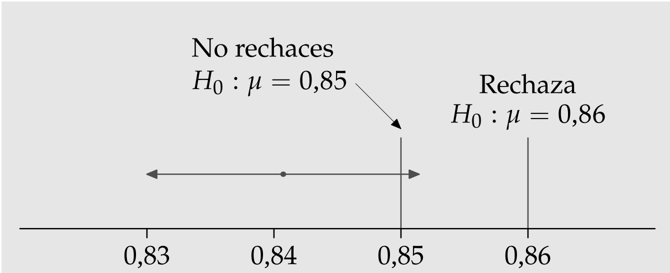


Figura 6.17. Los valores de μ que se hallan fuera del intervalo de confianza del 99% se pueden rechazar a un nivel de significación del 1%. Los valores que caen dentro del intervalo no se pueden rechazar.

APLICA TUS CONOCIMIENTOS

6.39. Coeficientes de inteligencia. He aquí los resultados de la prueba IQ de 31 estudiantes de primero de bachillerato:[9]

114	100	104	89	102	91	114	114	103	105	
108	130	120	132	111	128	118	119	86	72	
111	103	74	112	107	103	98	96	112	112	93

Trata a las 31 chicas como una muestra aleatoria simple de todas las chicas de primero de bachillerato de tu ciudad. Supón que la desviación típica de la IQ de esta población es conocida y sea $\sigma = 15$.

(a) Calcula un intervalo de confianza del 95% para la media de los IQ μ de la población.

(b) ¿Existe evidencia significativa a un nivel del 5% de que la media de los IQ de la población es diferente de 100? Plantea las hipótesis y utiliza tu intervalo de confianza para responder a la pregunta sin hacer más cálculos.

[9]Véase la nota 2.

RESUMEN DE LA SECCIÓN 6.3

Las **pruebas de significación** valoran 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, éstos proporcionan 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. 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 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 de una muestra**,

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

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 6.3

6.40. Estudio de satisfacción en el trabajo. El estudio del ejemplo 6.10 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 abscisas 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 abscisas 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.

6.41. 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?

6.42. Ciertos experimentos sobre aprendizaje en animales miden el tiempo que tarda un ratón en hallar la salida de un laberinto. El tiempo medio, para un laberinto determinado, resulta ser de 18 segundos. Una investigadora cree que un ruido ambiental fuerte provocará que los ratones hallen más rápidamente la salida. La investigadora determina el tiempo que tardan 10 ratones en encontrar la salida en estas condiciones. ¿Cuáles son las hipótesis nula H_0 y alternativa H_a?

6.43. ¿Se ha añadido agua a la leche? Una central lechera compra leche a varios proveedores. La central sospecha que algunos ganaderos añaden agua a la leche para aumentar sus beneficios. El exceso de agua se puede detectar midiendo el punto de congelación de la leche. La temperatura de congelación de la leche natural varía normalmente, con una media $\mu = -0{,}545$ °C y una desviación típica $\sigma = 0{,}008$ °C. La adición de agua aumenta la temperatura de congelación y la acerca a 0 °C, el punto de congelación del agua. El director del laboratorio de la central lechera determina la temperatura de congelación de cinco lotes consecutivos de leche procedentes de un mismo proveedor. La media es $\bar{x} = -0{,}538$ °C. Estos resultados, ¿constituyen una buena evidencia de que el proveedor está añadiendo agua a la leche? Plantea las hipótesis, contrástalas, da el valor P y redacta tus conclusiones.

6.44. Ingresos de altos ejecutivos. Un estudio sobre el salario de altos ejecutivos examinó, en un año reciente, el aumento de 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

$$H_0 : \mu = 0 \quad \text{(no aumentó)}$$
$$H_a : \mu > 0 \quad \text{(aumentó)}$$

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?

6.45. 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.

6.46. Existen otros estadísticos z que todavía no hemos estudiado. Puedes utilizar la tabla C para valorar la significación de cualquier estadístico z. Un estudio compara sociedades constituidas por empresas estadounidenses y japonesas en las que la empresa estadounidense es mayor que su socia japonesa, con otras sociedades en las que la empresa estadounidense es menor. Una de las variables determinadas en este estudio es el rendimiento de las acciones de la empresa estadounidense. La hipótesis nula es que "no hay diferencias" entre las medias de las dos poblaciones. La hipótesis alternativa es de dos colas. El valor del estadístico de contraste es $z = -1{,}37$.

(a) Este resultado, ¿es significativo a un nivel del 5%?

(b) ¿Y a un nivel del 10%?

6.47. Utiliza la tabla C para hallar el valor P aproximado en la prueba del ejercicio 6.40 sin hacer cálculos de probabilidad. Es decir, halla en la tabla dos valores que contengan el valor P entre ellos.

6.48. Utiliza la tabla C para hallar el valor P aproximado en la prueba del ejercicio 6.44. Es decir, ¿entre qué dos valores de la tabla debe encontrarse el valor P?

6.49. ¿Entre qué valores de la tabla C se halla el valor P del resultado $z = -1{,}37$ del ejercicio 6.46? (Recuerda que H_a es de dos colas.) Calcula el valor P utilizando la tabla A y comprueba que se encuentre entre los valores que hallaste en la tabla C.

6.50. Las ventajas de las patentes. Las empresas pioneras, es decir, empresas que se encuentran entre las primeras en desarrollar nuevos productos o servicios, tienden a tener mayores cuotas de mercado que sus competidoras menos innovadoras. ¿Qué explica esta ventaja? He aquí un extracto de las conclusiones de un estudio de una muestra de 1.209 fabricantes de productos industriales.

La posesión de una patente, ¿puede explicar las mayores cuotas de mercado de las empresas pioneras? Sólo el 21% de las empresas pioneras declara un beneficio significativo procedente de un producto patentado o de un secreto comercial. Aunque la media de la cuota de mercado de estas empresas es superior en dos puntos a la de otras empresas líderes que carecen de patentes, el aumento no es estadísticamente significativo ($z = 1{,}13$*). Por tanto, al menos en los mercados industriales desarrollados, las patentes de productos y los secretos comerciales tienen poca relación con las mayores cuotas de mercado de las empresas líderes.*[10]

[10]William T. Robinson, "Sources of market pioneer advantages: the case of industrial goods industries", *Journal of Marketing Research*, 25, febrero 1988, págs. 87-94.

Halla el valor P del valor de z dado. Luego explica a alguien que no sepa estadística lo que significa "no es estadísticamente significativo". ¿Por qué concluye el autor que las patentes y los secretos comerciales no representan una ayuda, pese a que contribuyeron un 2% a aumentar la media de la cuota de mercado?

6.51. Esta chica, ¿aparenta menos de 25 años? La industria tabacalera ha adoptado un código voluntario por el que se exige que las modelos que aparecen en sus anuncios deben aparentar al menos 25 años. Algunos estudios han demostrado, sin embargo, que los consumidores creen que muchas de las modelos son más jóvenes. Aquí tienes un fragmento de un estudio que preguntó a varias personas si 12 marcas de cigarrillos utilizaban modelos que aparentaban tener determinadas edades.

El ANCOVA reveló que la variable marca *es altamente significativa* ($P < 0{,}001$)*, lo que indica que la edad media de las modelos percibida por los sujetos del estudio no es igual para las 12 marcas. Como se comentó anteriormente, ciertas marcas como Lucky Strike Lights, Kool Milds y Virginia Slims tendían a tener modelos más jóvenes ...* [11]

El ANCOVA es una técnica estadística avanzada, pero la significación y los valores P tienen su significado habitual. Explica a alguien que no sepa estadística qué significa "altamente significativo ($P < 0{,}001$)" y por qué este resultado constituye una buena evidencia a favor de la existencia de diferencias de edad de las modelos que anuncian estas marcas, a pesar de que los sujetos sólo vieron una muestra de anuncios.

6.52. Explica con un lenguaje sencillo por qué una prueba de significación que es significativa a un nivel del 1% tiene que ser siempre significativa a un nivel del 5%.

6.53. Al explicar el sentido de "estadísticamente significativo a un nivel $\alpha = 0{,}05$" un estudiante dice: "esto quiere decir que la probabilidad de que la hipótesis nula sea cierta es menor de 0,05". ¿Es correcta esta explicación? ¿Por qué sí o por qué no?

[11] Michael B. Maziz *et al.*, "Perceived age and attractiveness of models in cigarette advertisement", *Journal of Marketing*, 56, enero 1992, págs. 22-37.

6.4 Utilización de las pruebas de significación

Las pruebas de significación se utilizan profusamente en las presentaciones de los resultados de trabajos de investigación en muchos campos de la ciencia y la tecnología. Para los nuevos productos farmacéuticos, la legislación exige que se demuestre su efectividad y su inocuidad. Los tribunales se interesan por la significación estadística en querellas por discriminación. Los investigadores de mercados quieren saber si una nueva campaña de publicidad será significativamente mejor que otra de efectos ya contrastados. Los investigadores médicos quieren saber si las nuevas terapias son significativamente mejores que las existentes. En cada una de todas estas situaciones, las pruebas de significación son apreciadas porque permiten detectar un efecto que es poco probable que ocurra únicamente por azar.

Llevar a cabo una prueba de significación resulta a menudo bastante sencillo, especialmente si calculas el valor P, sin esfuerzo, utilizando una calculadora o un programa estadístico. Utilizar acertadamente las pruebas de significación no es, en cambio, tan sencillo. Conviene, por lo tanto, hacer algunas consideraciones que debes tener presente cuando utilices o interpretes alguna prueba de significación.

6.4.1 ¿Cuál tiene que ser el valor P?

El objetivo de las pruebas de significación consiste en dejar bien claro el grado de evidencia proporcionado por la muestra en contra de la hipótesis nula. Esta evidencia se expresa mediante el valor P. ¿Cómo de pequeño tienen que ser el valor P para que exista evidencia clara en contra de la hipótesis nula? Esto depende principalmente de dos circunstancias:

- *¿Es plausible H_0?* Si H_0 representa una hipótesis en la que la gente que tienes que convencer ha creído durante años, se necesitará una evidencia fuerte (un valor P pequeño) para persuadirlos.
- *¿Cuáles son las consecuencias de rechazar H_0?* Si rechazar H_0 a favor de H_a significa, por ejemplo, un cambio de diseño que suponga un gasto económico importante, tienes que estar muy seguro de que el nuevo diseño hará aumentar las ventas.

Estos criterios son algo subjetivos. Personas distintas pueden querer utilizar distintos niveles de significación. Es, por tanto, mejor dar el valor P, para que cada cual pueda valorar según su propio criterio si la evidencia es suficientemente fuerte.

A menudo los usuarios de la estadística han utilizado niveles de significación del 10%, 5% y del 1%. Por ejemplo, los tribunales de justicia han tenido la tendencia a aceptar el 5% como estándar en los casos de discriminación.[12] Estos valores de referencia son un reflejo de la época en la que la utilización de valores críticos en vez de programas estadísticos era habitual en la estadística aplicada. El nivel de significación del 5% ($\alpha = 0{,}05$) ha sido el más utilizado. ***En cualquier caso, siempre tienes que tener bien presente que no existe ninguna frontera entre valores P "significativos" y "no significativos"; lo que sí ocurre es que a medida que el valor P disminuye, aumenta la evidencia en contra de*** H_0. A efectos prácticos no hay ninguna diferencia entre un valor P igual a 0,049 o uno igual a 0,051. No tiene sentido considerar $\alpha = 0{,}05$ como el valor de referencia universal para determinar lo que es significativo.

APLICA TUS CONOCIMIENTOS

6.54. ¿Es significativo? Supón que sin una preparación especial los resultados de la prueba SATM (Scholastic Assessment Test Mathematics) varían normalmente con $\mu = 475$ y $\sigma = 100$. Cien estudiantes pasan por un curso de preparación diseñado para mejorar sus resultados en la prueba SATM. Lleva a cabo una prueba de significación

$$H_0 : \mu = 475$$
$$H_a : \mu > 475$$

para contrastar en cada una de las siguientes situaciones:

(a) La media de los resultados de los estudiantes es $\bar{x} = 491{,}4$. Este resultado, ¿es significativo a un nivel del 5%?

(b) La media de los resultados es $\bar{x} = 491{,}5$. Este resultado, ¿es significativo a un nivel del 5%?

La diferencia entre los resultados en (a) y (b) no tiene importancia. No trates a $\alpha = 0{,}05$ como un número sagrado.

[12] D. H. Kaye, "Is proof of statistical significance relevant?" *Washington Law Review*, 61, 1986, págs. 1.333-1.365.

6.4.2 Significación estadística y significación práctica

Cuando se puede rechazar la hipótesis nula (de "ausencia de efecto" o de "no diferencia") a los niveles de significación habituales de $\alpha = 0{,}05$ o $\alpha = 0{,}01$, es que hay una buena evidencia a favor de que existe un efecto. De todas formas, éste puede ser muy pequeño. Cuando se dispone de muestras grandes, incluso las desviaciones muy pequeñas de la hipótesis nula pueden ser significativas.

EJEMPLO 6.18. Es significativo, ¿y qué?

Queremos contrastar la hipótesis de ausencia de correlación entre dos variables. Con 1.000 observaciones, una correlación de sólo $r = 0{,}08$ constituye una evidencia significativa a un nivel $\alpha = 0{,}01$ de que la correlación en la población no es cero sino que es positiva. Que el nivel de significación sea bajo no significa que la asociación sea fuerte, sólo indica que existe una fuerte evidencia de que existe alguna asociación. La verdadera correlación poblacional probablemente tome un valor cercano al valor muestral observado, $r = 0{,}08$. Podríamos, por tanto, concluir de forma correcta que, a efectos prácticos, se puede ignorar la asociación entre las dos variables, a pesar de que estamos seguros (a un nivel del 1%) de que la correlación es positiva. ■

El ejercicio 6.55 demuestra en detalle cómo al aumentar el tamaño de la muestra, disminuye el valor P. Recuerda la frase: "**la significación estadística no es lo mismo que la significación práctica".**

Un remedio para evitar dar demasiada importancia a la significación estadística es prestar tanta atención a los datos como al valor P. Representa gráficamente tus datos y examínalos cuidadosamente. ¿Existen observaciones atípicas u otras desviaciones? La existencia de unas pocas observaciones atípicas puede causar resultados altamente significativos si aplicas a ciegas las pruebas de significación. Las observaciones atípicas también pueden destruir la significación de unos datos que de otro modo darían resultados significativos. El uso imprudente de la estadística, alimentando los ordenadores con datos que no han sido sometidos a un análisis exploratorio previo, desemboca, a menudo, en afirmaciones absurdas. ¿Se puede ver en tus gráficos el efecto que estás buscando? Si no se ve, pregúntate si este efecto es suficientemente grande como para tener un interés práctico importante. Por regla general es prudente calcular un intervalo de confianza para el parámetro en el que estás interesado. Dar un intervalo de confianza es, de hecho, una forma de estimar la importancia del efecto, y no sólo una respuesta a la pregunta de si el efecto es demasiado grande para que ocurra sólo por azar. Los intervalos de confianza no se utilizan tan a menudo como se debería, mientras que las pruebas de significación quizás se utilizan demasiado.

APLICA TUS CONOCIMIENTOS

6.55. Preparación para la prueba SAT. Supongamos que los resultados de la prueba SATM cuando los estudiantes no reciben una preparación especial se distribuyen normalmente con $\mu = 475$ y $\sigma = 100$. Supón también que una preparación especial pueda cambiar μ sin modificar σ. Un aumento en la puntuación de 475 a 478 no influye sobre la admisión de un alumno en la universidad, pero este cambio puede ser estadísticamente muy significativo. Para comprobarlo, calcula el valor P para el contraste

$$H_0 : \mu = 475$$
$$H_a : \mu > 475$$

en cada una de las siguientes situaciones:

(a) Cien estudiantes asisten a unas clases especiales de preparación para la prueba SATM. Su media en esta prueba es $\bar{x} = 478$.

(b) Al año siguiente, los estudiantes que asisten a las clases son 1.000. La media de estos estudiantes en la prueba SATM es $\bar{x} = 478$.

(c) Después de una fuerte campaña de publicidad, son 10.000 los estudiantes que asisten a las clases de preparación. La media de los resultados de estos alumnos en la prueba SATM se mantiene en $\bar{x} = 478$.

6.56. Calcula un intervalo de confianza del 99% para la media μ de los resultados de la prueba SATM en cada apartado del ejercicio anterior. Para muestras grandes, el intervalo de confianza nos dice, "sí, el resultado medio supera los 475 puntos después de haber asistido a las clases, pero por muy poco".

6.4.3 La inferencia estadística no es válida para cualquier conjunto de datos

Vamos a insistir de nuevo en que los datos procedentes de encuestas o experimentos mal diseñados a menudo dan resultados carentes de validez. La inferencia estadística no puede corregir los defectos de un mal diseño. Cada prueba es válida sólo en determinadas circunstancias, siendo de especial importancia la obtención correcta de los datos. La prueba z, por ejemplo, debería ir acompañada del mismo listado de advertencias que vimos anteriormente en este capítulo, referidas a los intervalos de confianza. Advertencias similares son también aplicables a otras pruebas que estudiaremos.

EJEMPLO 6.19. Ligazón de arterias mamarias

Se produce una angina de pecho cuando no llega suficiente sangre al corazón. Quizá se podrían aliviar las anginas inutilizando las arterias mamarias para obligar al cuerpo a desarrollar otras rutas que suministren sangre al corazón. Unos cirujanos ensayaron este procedimiento, conocido como ligazón de las arterias mamarias. Los pacientes mostraron una reducción estadísticamente significativa en la incidencia de anginas de pecho.

La inferencia estadística nos dice que no sólo ha actuado el azar, pero no nos dice qué es lo que ha actuado. El experimento de la ligazón de las arterias mamarias fue incontrolado, de manera que la reducción de las anginas de pecho podía ser debido al efecto placebo. Posteriormente, un experimento comparativo aleatorizado mostró que la ligazón no era más efectiva que el placebo. Los cirujanos abandonaron inmediatamente esta práctica. ■

Las pruebas de significación y los intervalos de confianza se basan en las leyes de la probabilidad. La aleatorización en el muestreo y en los experimentos garantiza que se puedan aplicar dichas leyes. Sin embargo, a menudo hemos de analizar datos que no se han obtenido a partir de muestras o experimentos aleatorizados. En tales casos, para poder realizar la inferencia estadística, tenemos que poder describir la distribución de los datos en términos de probabilidad. Los diámetros de una serie de perforaciones hechas en motores de coche en una cadena de producción, por ejemplo, pueden comportarse igual que una muestra aleatoria de una distribución normal. Podemos comprobar su distribución de probabilidad examinando los datos. Si la distribución de los datos es normal, podemos aplicar las fórmulas de este capítulo para hacer inferencia sobre el diámetro medio μ de las perforaciones. En definitiva, pregunta siempre cómo se obtuvieron los datos y no te dejes impresionar demasiado por los valores P hasta que no estés completamente seguro de que se puedan utilizar, con garantías, las pruebas de significación.

APLICA TUS CONOCIMIENTOS

6.57. Preguntas a telespectadores. Una emisora de televisión local plantea preguntas que los telespectadores contestan por teléfono. La pregunta de hoy hace referencia a un proyecto de ley que prohibe fumar en los restaurantes. De las 2.372 llamadas recibidas, 1.921 se oponen a la nueva ley. La emisora, siguiendo las prácticas estadísticas habituales, hace una afirmación sobre la confianza de los resultados: "el 81% de la muestra de la encuesta del Canal 13 se opone a que se prohiba fumar en los restaurantes. Podemos tener una confianza del 95% en que la proporción de todos los televidentes que se oponen a la ley no difiere en más del 1,6% del resultado muestral". ¿Está justificada la conclusión de la emisora? Justifica tu respuesta.

6.4.4 Cuidado con los análisis múltiples

Las pruebas de significación parecen indicar que se ha encontrado el efecto buscado. Pero sólo tienen sentido si se ha decidido previamente cuál es el efecto que se está buscando, y si se ha diseñado cuidadosamente la forma de identificarlo. En otros casos, las pruebas de significación pueden tener poco sentido.

EJEMPLO 6.20. Trayectoria de los ejecutivos en formación

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. ■

Mejor que contrastar todas las variables de los ejecutivos en formación, podrías buscar la variable con más diferencias entre los ejecutivos que llegan a cargos de responsabilidad de los que finalmente abandonan la empresa y luego contrastar si esta diferencia es significativa. Esta manera de proceder tampoco es correcta. El valor P presupone que tú ya sabías qué comparaciones querías hacer antes de observar los datos. Es hacer trampas observar entre qué variables hay más diferencias y luego actuar como si ésta fuera la variable escogida antes de hacer la prueba.

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.

APLICA TUS CONOCIMIENTOS

6.58. Percepción extrasensorial. 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 aquellos casos en que se responde 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 DE LA SECCIÓN 6.4

No existe una regla universal que nos diga como tiene que ser de pequeño un valor P para ser convincente. Evita dar demasiada importancia a los niveles de significación tradicionales como por ejemplo $\alpha = 0{,}05$.

Efectos muy pequeños pueden ser muy significativos (valores P pequeños), especialmente cuando una 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 intervalos de confianza para estimar los verdaderos valores de los parámetros.

Por otro lado, la falta de significación no implica 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 6.4

6.59. ¿Para qué sirve la significación? ¿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?

6.60. Detectores de radares y velocidad. Unos investigadores observaron la velocidad de los automóviles que circulaban por una autopista (con límite de velocidad de 120 km por hora) antes y después de hacer funcionar un radar. Los investigadores compararon la velocidad de los coches que tienen detectores de radar con la velocidad de los coches sin detectores. He aquí las velocidades medias (en kilómetros por hora) observadas para 22 coches con detectores de radar y 46 sin ellos.[13]

[13]N. Teed, K. L. Adrian y R. Knoblouch, "The duration of speed reductions attributable to radar detectors", *Accident Analysis and Prevention*, 25, 1991, págs. 131-137.

	¿Detector de radar?	
	Sí	**No**
Sin el radar	113	109
Con el radar	95	108

Dice el trabajo: "Los vehículos que tenían detector de radar iban más deprisa ($P < 0{,}01$) que los vehículos que no lo tenían antes de estar expuestos al radar y significativamente más lentos inmediatamente después ($P < 0{,}0001$)".

(a) Explica por qué estos valores P constituyen una buena evidencia de que los conductores con detectores de radar se comportan de forma distinta.

(b) A pesar de que $P < 0{,}01$, antes de conectarse los radares, la diferencia entre las velocidades medias de los conductores con detectores de radar y sin ellos es pequeña. Explica en un lenguaje sencillo cómo una diferencia tan pequeña puede ser estadísticamente significativa.

6.61. 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.

6.62. ¿Qué distingue a los esquizofrénicos? En un experimento, unos psicólogos midieron 77 variables de una muestra de esquizofrénicos 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%?

6.5 Tipos de error y potencia*

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

*Esta sección más avanzada introduce nuevos conceptos relacionados con las pruebas estadísticas. Esta sección no es necesaria para leer el resto del libro.

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 la hipótesis nula.

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 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.

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. 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 describir el comportamiento de las pruebas de significación desde otra óptica.

Controles de calidad

6.5.1 Errores tipo I y errores tipo II

En los controles de calidad tenemos que decidir entre

$$H_0 \quad : \quad \text{el envío de cojinetes cumple los estándares de calidad}$$

$$H_a \quad : \quad \text{el envío de cojinetes no cumple los estándares de calidad}$$

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 o 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 TIPO I Y ERRORES TIPO II

Si rechazamos H_0 cuando en realidad H_0 es cierta, cometemos un **error tipo I.**

Si no rechazamos H_0 cuando en realidad H_a es cierta, cometemos un **error tipo II.**

La figura 6.18 muestra las cuatro situaciones posibles. Cuando H_0 es cierta, nuestra decisión es correcta si aceptamos H_0 o cometemos un error tipo I si rechazamos H_0. Cuando H_a es cierta, nuestra decisión es correcta o cometemos un error 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 I

Figura 6.18. Los dos tipos de error en el contraste de hipótesis.

6.5.2 Probabilidades de error

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 rechazar H_0 significa decidir que H_0 es cierta. Podemos, pues, describir el resultado de una prueba mediante las probabilidades de los errores tipo I y tipo II. 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?"

EJEMPLO 6.21. Este envío de cojinetes, ¿es correcto?

Se supone que el diámetro medio de un determinado tipo de cojinetes es de 2,000 centímetros (cm). El diámetro de los cojinetes varía según una distribución normal con una desviación típica $\sigma = 0{,}010$ cm. Cuando llega un envío, el comprador toma una muestra aleatoria simple de 5 cojinetes 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 tipo I significa rechazar H_0 cuando en realidad $\mu = 2$.

¿Qué podemos decir sobre los errores 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 debería rechazar un envío de cojinetes cuando su diámetro medio sea igual a 2,015 cm. Por tanto, se cometerá un error tipo II cuando se acepte H_0 siendo en realidad $\mu = 2{,}015$.

La figura 6.19 muestra cómo se obtienen las dos probabilidades de error a partir de las dos distribuciones 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 tipo I. Cuando $\mu = 2{,}015$, H_a es cierta y aceptar H_0 significa cometer un error tipo II. A continuación calcularemos las probabilidades de cada uno de estos tipos de error. ■

La probabilidad de un error 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 valores z tan extremos ocurrirán con una probabilidad igual a 0,05 cuando H_0 sea cierta.

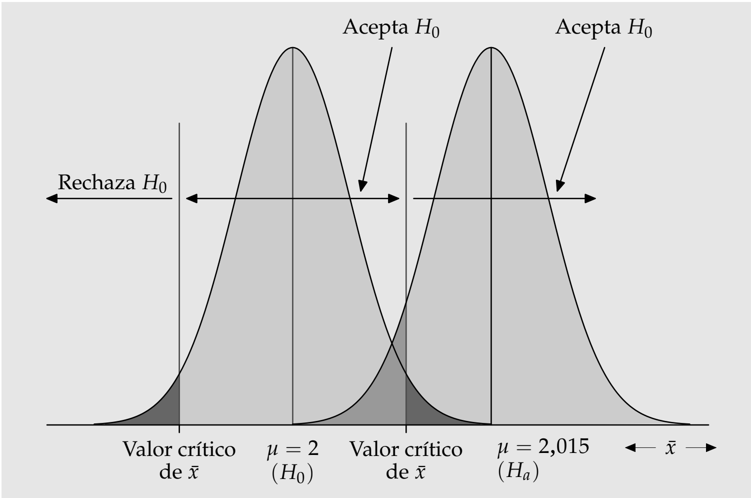


Figura 6.19. Las probabilidades de los dos tipos de error del ejemplo 6.21. La probabilidad del error tipo I (*el área sombreada más clara*) es la probabilidad de rechazar $H_0 : \mu = 2$ cuando en realidad $\mu = 2$. La probabilidad del error tipo II (*área sombreada más oscura*) es la probabilidad de aceptar H_0 cuando en realidad $\mu = 2{,}015$.

NIVEL DE SIGNIFICACIÓN Y ERROR TIPO I

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

La probabilidad de un error tipo II para la alternativa $\mu = 2{,}015$ del ejemplo 6.21 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 $\mu = 2$. Veamos el cálculo del error tipo II.

EJEMPLO 6.22. Cálculo del error tipo II

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

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

que es lo mismo que,

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

o que, una vez realizado el cálculo,

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

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.

$$\begin{aligned} P(\text{error tipo II}) &= P(1{,}9912 \leq \bar{x} \leq 2{,}0088) \\ &= P\left(\frac{1{,}9912 - 2{,}015}{0{,}01/\sqrt{5}} \leq \frac{\bar{x} - 2{,}015}{0{,}01/\sqrt{5}} \leq \frac{2{,}0088 - 2{,}015}{0{,}01/\sqrt{5}}\right) \\ &= P(-5{,}32 \leq Z \leq -1{,}39) \\ &= 0{,}0823 \end{aligned}$$

La figura 6.20 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 $\mu = 2$). Esta prueba 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.

APLICA TUS CONOCIMIENTOS

6.63. Tu empresa comercializa un programa de diagnóstico médico informatizado. El programa comprueba los resultados de las pruebas médicas rutinarias

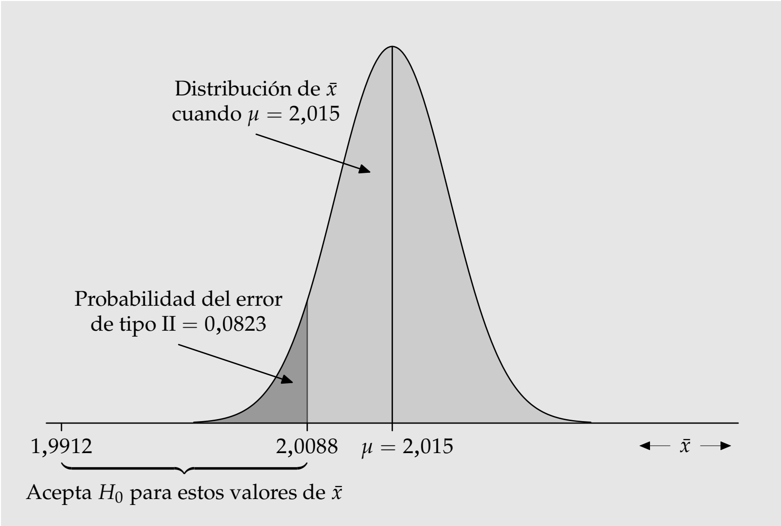


Figura 6.20. La probabilidad del error tipo II del ejemplo 6.22. Ésta es la probabilidad de que la prueba acepte H_0 cuando la hipótesis alternativa es cierta.

(presión sanguínea, análisis de sangre, etc.) y 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.)

6.64. 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 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 tipo II cuando $\mu = 270$.

6.65. 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 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 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 tipo II cuando $\mu = 1$.

6.5.3 Potencia

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

POTENCIA

La probabilidad de que una prueba con un nivel de significación predeterminado α rechace H_0 cuando el parámetro toma un cierto valor alternativo se llama **potencia** de la prueba con esta alternativa.

La potencia de una prueba con cualquier alternativa es 1 menos la probabilidad del error tipo II para esta alternativa.

Los cálculos de la potencia son esencialmente los mismos que los cálculos de probabilidad de los errores tipo II. En el ejemplo 6.22, 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 fuera 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 confirma que la potencia sea la adecuada. Si la potencia es demasiado baja, una muestra mayor la aumentará 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 $\alpha = 0{,}05$, incluso si se tiene la intención de dar el valor P.

APLICA TUS CONOCIMIENTOS

6.66. El fabricante de refrescos del ejercicio 6.8 considera que la pérdida de dulzor sería inaceptable si la media de las respuestas de todos los catadores fuera $\mu = 1{,}1$. ¿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 con la alternativa $\mu = 1{,}1$. Esta potencia es la probabilidad de que la prueba rechace H_0 cuando $\mu = 1{,}1$ sea cierta. El método de cálculo es similar al del error tipo II.

(a) Paso 1. *Escribe el procedimiento para rechazar H_0 en términos de $\bar{x}$*. Sabemos que $\sigma = 1$, por lo que la prueba rechaza H_0 a un nivel $\alpha = 0{,}05$ 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 $\mu = 1{,}1$ para hallar la probabilidad de que $\bar{x}$ tome un valor que lleve a rechazar H_0.

6.67. El ejercicio 6.36 hace referencia a una prueba sobre el contenido medio de botellas de cola. 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 $\sigma = 3$. Una prueba de significación del 5% rechaza H_0 si $z \leq -1{,}645$, 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 con la alternativa $\mu = 299$.

(b) Halla la potencia de la prueba con la alternativa $\mu = 295$.

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

6.68. 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 con $\mu = 299$ cuando $n = 25$.
(b) Halla la potencia de la prueba con $\mu = 299$ cuando $n = 100$.

RESUMEN DE LA SECCIÓN 6.5

Las **pruebas con un nivel de determinación α predeterminado** se utilizan algunas veces para decidir si se acepta H_0 o H_a.

Describimos el comportamiento de una prueba con un nivel de determinación predeterminado dando las probabilidades de dos tipos de error. Cometemos un **error tipo I** si rechazamos H_0 cuando en realidad H_0 es cierta. Cometemos un **error tipo II** si no rechazamos 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 con una determinada alternativa es la probabilidad de que la prueba rechace H_0 cuando esa 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 tipo I, y la potencia con una determinada alternativa es igual a 1 menos la probabilidad del error tipo II de esa alternativa.

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

EJERCICIOS DE LA SECCIÓN 6.5

6.69. Los cálculos de potencia de las pruebas de dos colas siguen la misma pauta que los cálculos de potencia de las pruebas de una cola. El ejemplo 6.16 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 con la alternativa $\mu = 0{,}845$.

(a) La prueba del ejemplo 6.16 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 tipo II cuando $\mu = 0{,}845$?

6.70. En el ejemplo 6.13 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 en caso de existir una diferencia importante, la prueba utilizada detectaría esta diferencia. 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 con la alternativa $\mu = 134$.

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

(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)?

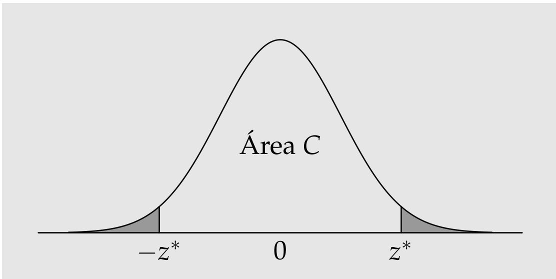
6.71. En el ejercicio 6.67 hallaste la potencia de una prueba con la alternativa $\mu = 295$. Utiliza el resultado de ese ejercicio para hallar las probabilidades de los errores tipo I y tipo II para esa prueba y esa alternativa.

6.72. En el ejercicio 6.64 hallaste las probabilidades de los dos tipos de error de la prueba $H_0 : \mu = 275$, con la alternativa $\mu = 270$. Utiliza el resultado de ese ejercicio para dar la potencia de la prueba con la alternativa $\mu = 270$.

6.73. El mercado de valores, ¿es eficiente? Lees un artículo en un periódico económico que comenta la "hipótesis del mercado eficiente" para explicar el comportamiento de las cotizaciones de los valores bursátiles. El autor admite que la mayoría de contrastes de esta hipótesis no han hallado evidencia significativa en

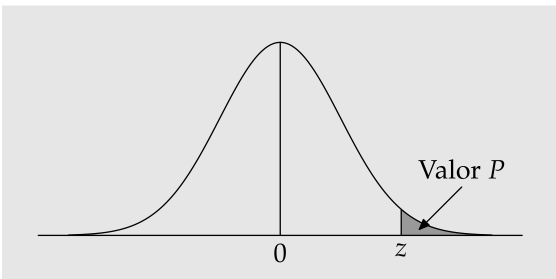
Concepto de intervalo de confianza

$$
\left.
\begin{array}{lcr}
\textbf{Población} & \xrightarrow{\text{Muestra aleatoria simple de tamaño } n} & \bar{x} \pm z^* \dfrac{\sigma}{\sqrt{n}} \\
\mu = ? & \xrightarrow{\text{Muestra aleatoria simple de tamaño } n} & \bar{x} \pm z^* \dfrac{\sigma}{\sqrt{n}} \\
\sigma \text{ conocida} & \xrightarrow{\text{Muestra aleatoria simple de tamaño } n} & \bar{x} \pm z^* \dfrac{\sigma}{\sqrt{n}} \\
 & \vdots & \vdots
\end{array}
\right\} \text{Una proporción } C \text{ de estos intervalos capturan la verdadera } \mu
$$



Concepto de prueba de significación

$$
\left.
\begin{array}{lcr}
\textbf{Población} & \xrightarrow{\text{Muestra aleatoria simple de tamaño } n} & z = \dfrac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} \\
H_0 : \mu = \mu_0 & & \\
H_a : \mu > \mu_0 & \xrightarrow{\text{Muestra aleatoria simple de tamaño } n} & z = \dfrac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} \\
\sigma \text{ conocida} & \xrightarrow{\text{Muestra aleatoria simple de tamaño } n} & z = \dfrac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} \\
 & \vdots & \vdots
\end{array}
\right\}
$$



su contra. Pero insinúa que esto es debido a que las pruebas utilizadas tienen poca potencia. "La impresión generalizada de que existe una fuerte evidencia a favor de la eficiencia del mercado puede ser debida simplemente a la baja potencia de muchas pruebas estadísticas".[14]

Explica con lenguaje sencillo por qué las pruebas que tienen poca potencia a menudo no consiguen proporcionar evidencia en contra de una hipótesis incluso cuando ésta es realmente falsa.

REPASO DEL CAPÍTULO 6

La inferencia estadística saca conclusiones sobre una población a partir de una muestra de datos y utiliza la probabilidad para indicar la fiabilidad de las conclusiones. Un intervalo de confianza estima un parámetro desconocido. Una prueba de significación muestra la fuerza de la evidencia a favor de una afirmación sobre un parámetro.

Las probabilidades de las pruebas de significación y de los intervalos de confianza nos dicen lo que ocurriría si utilizáramos las fórmulas de los intervalos o de las pruebas de significación muchísimas veces. Un nivel de confianza es la probabilidad de que la fórmula de un intervalo de confianza dé realmente un intervalo que contenga el parámetro desconocido. Un intervalo de confianza del 95% da un resultado correcto en un 95% de las veces cuando lo utilizamos de forma repetida. Un valor P es la probabilidad de que la prueba dé un resultado al menos tan extremo como el resultado observado si la hipótesis nula fuera realmente cierta. Esto es, el valor P nos indica si el resultado obtenido es sorprendente. Los resultados muy sorprendentes (con valores P pequeños) constituyen una buena evidencia a favor de que la hipótesis nula no es cierta.

Las figuras de la página anterior presentan de forma gráfica las ideas básicas sobre los intervalos de confianza y las pruebas de significación. Estas ideas son la base de los próximos capítulos del libro. Hablaremos extensamente sobre muchos métodos estadísticos y también de su aplicación práctica. En cada caso, los razonamientos básicos sobre los intervalos de confianza y las pruebas de significación son los mismos. He aquí lo más importante que tienes que ser capaz de hacer después de estudiar este capítulo.

[14] Robert J. Schiller, "The volatility of stock market prices", *Science*, 235, 1987, págs. 33-36.

A. INTERVALOS DE CONFIANZA

1. Describir con un lenguaje sencillo cuál es el significado de "un intervalo de confianza del 95%" o de otras expresiones sobre los intervalos de confianza que aparecen en los informes estadísticos.
2. Calcular un intervalo de confianza para la media μ de una población normal de desviación típica σ conocida, utilizando la fórmula $\bar{x} \pm z^* \frac{\sigma}{\sqrt{n}}$.
3. Identificar cuándo puedes utilizar con seguridad la fórmula anterior y cuándo el diseño muestral o el tamaño demasiado pequeño de una muestra de una población muy asimétrica la hacen inexacta. Comprender que el error de estimación no incluye los efectos de la falta de cobertura, de la no-respuesta u otras dificultades de orden práctico.
4. Comprender cómo cambia el error de estimación de un intervalo de confianza al variar el tamaño de la muestra y el nivel de confianza C.
5. Hallar el tamaño de la muestra que se requiere para obtener un intervalo de confianza con un determinado error de estimación m cuando se ha dado el nivel de confianza y otras informaciones.

B. PRUEBAS DE SIGNIFICACIÓN

1. Plantear las hipótesis nula y alternativa cuando el parámetro que hay que contrastar es la media poblacional μ.
2. Explicar con un lenguaje sencillo el significado del valor P cuando te dan su valor numérico de una prueba.
3. Calcular el estadístico z y el valor P para pruebas de una y de dos colas sobre la media μ de una población normal.
4. Valorar la significación estadística a niveles estándar α, comparando P con α o comparando z con los valores críticos de la distribución normal estandarizada.
5. Saber que las pruebas de significación no miden la magnitud o la importancia de un efecto.
6. Saber cuándo puedes utilizar el estadístico z y cuándo el diseño de la obtención de los datos o el tamaño demasiado pequeño de una muestra de una población asimétrica lo hacen inapropiado.

EJERCICIOS DE REPASO DEL CAPÍTULO 6

6.74. El olor del vino. Los compuestos sulfurosos son la causa de ciertos olores desagradables en el vino. Por este motivo, los enólogos quieren conocer el umbral de percepción del olor, es decir, la concentración más baja de un compuesto que el olfato humano puede detectar. El umbral de percepción del olor del sulfato de dimetilo para los catadores experimentados es aproximadamente de 25 microgramos por litro de vino (μg/l). Sin embargo, los olfatos menos entendidos de los consumidores pueden ser menos sensibles. He aquí los umbrales de percepción del sulfato de dimetilo de 10 consumidores poco experimentados:

31 31 43 36 23 34 32 30 20 24

Supón que la desviación típica del umbral de percepción del olor de los olfatos poco experimentados es $\sigma = 7$ μg/l.

(a) Dibuja un diagrama de tallos para verificar que la distribución es aproximadamente simétrica sin observaciones atípicas. (Más datos confirman que no hay desviaciones sistemáticas de la normalidad.)

(b) Calcula un intervalo de confianza del 95% para la media del umbral de percepción del sulfato de dimetilo de todos los consumidores.

(c) ¿Estás convencido de que la media del umbral de percepción del olor de los consumidores es mayor que el umbral de 25 μg/l? Lleva a cabo una prueba de significación que justifique tu respuesta.

6.75. Celulosa de la alfalfa. Un ingeniero agrónomo examina el contenido de celulosa de una determinada variedad de alfalfa. Supón que el contenido de celulosa de la población tiene una desviación típica $\sigma = 8$ mg/g. Una muestra de 15 cortes de alfalfa tiene un contenido medio de celulosa $\bar{x} = 145$ mg/g.

(a) Calcula un intervalo de confianza del 90% para el contenido medio de celulosa de la población.

(b) Un estudio afirma que el contenido medio de celulosa es $\mu = 140$ mg/g, pero el ingeniero agrónomo cree que la media es mayor que ese valor. Plantea H_0 y H_a y lleva a cabo una prueba de significación para ver si los nuevos datos corroboran esta impresión.

(c) Los procedimientos estadísticos utilizados en (a) y (b) son válidos cuando se cumplen varios supuestos. ¿Qué supuestos son?

6.76. Deficiencias de hierro en bebés. Unos investigadores, que estudian las deficiencias de hierro en bebés, examinaron niños que seguían distintas dietas alimenticias. Un grupo de 26 bebés se alimentaba con leche materna. A los 6 meses, estos niños tenían un nivel medio de hemoglobina de $\bar{x} = 12{,}9$ gramos por 100 mililitros (g/100 ml) de sangre. Supón que la desviación típica de la población es $\sigma = 1{,}6$ g/100 ml. Calcula un intervalo de confianza del 95% para el nivel medio de hemoglobina de los niños alimentados con leche materna. ¿Qué supuestos (aparte del poco realista de que conocemos σ) exige el método que utilizaste para calcular el intervalo de confianza?

6.77. Inversión en letras del tesoro. La compra de letras del tesoro es una inversión segura, pero, ¿es una inversión rentable? He aquí datos sobre el rendimiento (en porcentaje) de las letras del tesoro de 1970 a 1996:

Año	1970	1971	1972	1973	1974	1975	1976	1977	1978	1979
Rendimiento	6,45	4,37	4,17	7,20	8,00	5,89	5,06	5,43	7,46	10,56
Año	1980	1981	1982	1983	1984	1985	1986	1987	1988	1989
Rendimiento	12,18	14,71	10,84	8,98	9,89	7,65	6,10	5,89	6,95	8,43
Año	1990	1991	1992	1993	1994	1995	1996			
Rendimiento	7,72	5,46	3,50	3,04	4,37	5,60	5,13			

(a) Dibuja un histograma con estos datos, utiliza barras que tengan una anchura de 2 puntos porcentuales. ¿Qué clase de desviación de la normalidad observas? Gracias al teorema del límite central, puedes sin embargo tratar a $\bar{x}$ como aproximadamente normal.

(b) Supón que podemos considerar estos 27 años como una muestra aleatoria de los rendimientos de las letras del tesoro. Calcula un intervalo de confianza del 90% para la media de los rendimientos a lo largo de muchos años. (Supón que conoces que la desviación típica de todos los rendimientos es $\sigma = 2{,}75\%$.)

(c) La tasa de inflación media durante estos años es del 5,5%. ¿Estás seguro de que el rendimiento medio de las letras del tesoro es superior al 5,5%? Plantea las hipótesis, calcula el estadístico de contraste y halla el valor P.

(d) Dibuja un diagrama temporal con estos datos. Se observan ciclos de subidas y bajadas en los rendimientos que siguen los ciclos de los tipos de interés. El diagrama temporal deja claro que los rendimientos de los sucesivos años están fuertemente correlacionados; por tanto, no es correcto tratar estos datos como una muestra aleatoria simple. Siempre tienes que comprobar que no existe este *tipo de correlaciones* con datos obtenidos a lo largo del tiempo.

6.78. Ingresos de trabajadores a tiempo completo. Los organismos oficiales, a menudo utilizan los intervalos de confianza del 90% en sus informes. Uno de ellos proporciona un intervalo de confianza del 90% para la media de los ingresos de los trabajadores a tiempo completo en 1997 como de 33.148 a 34.200 €. Este intervalo se calculó mediante métodos avanzados de la encuesta de la población activa, una muestra aleatoria en etapas múltiples de unos 50.000 hogares.

(a) Un intervalo de confianza del 95%, ¿sería más amplio o más estrecho? Justifica tu respuesta.

(b) La hipótesis nula de que la media de los ingresos de los trabajadores a tiempo completo en 1997 era de 33.000 €, ¿se rechazaría a un nivel de significación del 1% si se plantea una alternativa de dos colas? ¿Qué podríamos decir si la hipótesis nula planteara que la media era de 34.000 €?

6.79. ¿Por qué son mejores las muestras grandes? En estadística se prefieren muestras grandes. Describe brevemente la consecuencia de aumentar el tamaño de una muestra (o el número de sujetos en un experimento) en cada uno de los siguientes casos:

(a) El error de estimación de un intervalo de confianza del 95%.

(b) El valor P de una prueba, cuando H_0 es falsa y todas las características de la población se mantienen constantes cuando aumenta n.

(c) (Optativo) La potencia de una prueba con un nivel α predeterminado, cuando α, la hipótesis alternativa y todos las características de la población permanecen constantes.

6.80. Una ruleta tiene 18 casillas rojas de un total de 38. Observas los resultados y anotas el número de veces que aparece el rojo. Ahora quieres utilizar estos datos para contrastar si la probabilidad p de que aparezca rojo es la que tendría que ser si la ruleta estuviese bien construida. Plantea las hipótesis H_0 y H_a que contrastarás. (Describiremos la prueba adecuada para esta situación en el capítulo 8.)

6.81. Cuando se pregunta a un grupo de estudiantes que expliquen el significado de "el valor P era 0,03", un estudiante dice, "significa que sólo existe una probabilidad igual a 0,03 de que la hipótesis nula sea cierta". Esta explicación, ¿es correcta? Justifica tu respuesta.

6.82. Cuando se pregunta a otro estudiante por qué la significación estadística aparece tan frecuentemente en los trabajos de investigación, contesta, "porque decir que los resultados son significativos nos indica que no se pueden explicar fácilmente sólo por la variación debida al azar". ¿Crees que esta afirmación es esencialmente correcta? Justifica tu respuesta.

6.83. Madres que necesitan asistencia social. Un estudio compara dos grupos de madres con niños pequeños que necesitaron ayuda de los servicios de asistencia social hace dos años. Un grupo asistió voluntariamente a un programa gratuito de formación que se anunció en la prensa local. El otro no asistió al programa. El estudio halla una diferencia significativa ($P < 0{,}01$) entre las proporciones de madres de los dos grupos que todavía necesitan ayuda de los servicios de asistencia social. La diferencia no es sólo significativa sino que es bastante grande. El informe afirma, con una confianza del 95%, que el porcentaje de madres que todavía necesita ayuda del grupo que no asistió al programa de formación es un 21% $\pm$ 4% mayor que el porcentaje de madres del grupo que asistió al programa de formación. Se te pide que valores las conclusiones del informe.

(a) Explica con un lenguaje sencillo qué significa "una diferencia significativa ($P < 0{,}01$)".

(b) Explica de una manera clara y brevemente qué significa "una confianza del 95%".

(c) Este estudio, ¿constituye una buena evidencia a favor de exigir a las madres que necesitan asistencia social su participación en programas de formación? La participación en estos programas, ¿reduciría significativamente el porcentaje de madres que siguen necesitando asistencia social durante varios años?