

13 INFERENCIA ESTADÍSTICA. ESTIMACIÓN DE UNA PROPORCIÓN

Página 309

Resuelve

¿Cuántas caras cabe esperar?

- Repite el razonamiento anterior para averiguar *cuántas caras cabe esperar* si lanzamos 100 monedas y consideramos «casos raros» al 5 % de los casos extremos.

El intervalo característico correspondiente a una probabilidad del 95 % (consideramos «casos raros» al 5 % de los casos extremos) es:

$$50 \pm 1,96 \cdot 5 = (40,2; 59,8)$$

Esto significa que en el 95 % de los casos en que tiremos 100 monedas, el número de caras que obtendremos será mayor que 40 y menor que 60. Cualquier otro resultado será un «caso raro».

Un saco de alubias

Tenemos un saco con 10 000 alubias. De ellas, 9 500 son blancas y 500 son negras. Están bien mezcladas.

Extraemos 600 alubias.

¿Cuántas alubias negras cabe esperar que haya entre ellas?

- Resuelve el problema anterior considerando como «casos raros» solo al 1 % de los casos extremos. Para ello:

- Averigua la proporción, p , de alubias negras en el saco.
- Considera la distribución $B(600, p)$ y calcula su media $\mu = 600p$ y su desviación típica $\sigma = \sqrt{600 \cdot p(1-p)}$.
- Considera la distribución $N(\mu, \sigma)$ y halla su intervalo característico correspondiente a una probabilidad del 99 %.
- Decide, como consecuencia del resultado anterior, entre qué valores se encuentra el número de alubias negras que *cabe esperar*.

a) $p = \frac{500}{10\,000} = 0,05$

b) $\mu = 600 \cdot 0,05 = 30$; $\sigma = \sqrt{600 \cdot 0,05 \cdot 0,95} = \sqrt{28,5} \approx 5,34$

- c) El intervalo característico correspondiente a una probabilidad del 99 % es:

$$30 \pm 2,575 \cdot 5,34 = (16,25; 43,75)$$

- d) En el 99 % de los casos en que saquemos 600 judías de ese saco, el número de judías negras será mayor que 16 y menor que 44. Cualquier otro resultado será un «caso raro» (llamando «casos raros» a ese 1 % de casos extremos).

Peces en un pantano

Se desea estimar el número total de peces que hay en cierto pantano. Para ello, se procede del siguiente modo:

- Se cogen con una red una cierta cantidad de ellos, por ejemplo, 349, se marcan y se devuelven al pantano.
- Al cabo de varios días, se vuelve a coger otro montón.

Supongamos que en esta segunda pesca se han obtenido 514 peces, de los cuales hay 37 marcados.

- Con los datos anteriores, di cuántos peces crees que hay, aproximadamente, en el pantano.

La muestra tiene 514 peces, de los cuales hay 37 marcados. La proporción de peces marcados en la muestra es:

$$pr = \frac{37}{514} = 0,072$$

El valor de la proporción de peces marcados en el pantano es $pr = \frac{349}{N}$, donde N es el número total de peces.

Aunque este problema se resolverá de forma completa (mediante un intervalo de confianza) al terminar la unidad, podemos suponer que la proporción de peces marcados en la muestra y en el pantano será «aproximadamente» la misma; es decir:

$$\frac{37}{514} \approx \frac{349}{N} \rightarrow N \approx 4\,848,27 \rightarrow N \approx 4\,848 \text{ peces}$$

(Al considerar una probabilidad determinada, daremos un intervalo de confianza, obteniendo un resultado más preciso que este).

1 ► DISTRIBUCIÓN BINOMIAL. REPASO DE TÉCNICAS BÁSICAS PARA EL MUESTREO

Página 311

1 La variable x es binomial, con $n = 1\,200$ y $p = 0,008$.

a) Calcula la probabilidad de que x sea mayor que 100.

b) Halla el intervalo característico para una probabilidad del 95 %.

Como $np = 9,6 > 5$ y $nq > 5$, podemos aproximar mediante una normal de media $\mu = np = 9,6$ y desviación típica $\sigma = \sqrt{npq} = \sqrt{1\,200 \cdot 0,008 \cdot 0,992} = 3,09$.

Es decir:

$$x \text{ es } B(1\,200; 0,008) \rightarrow x' \text{ es } N(9,6; 3,09) \rightarrow z \text{ es } N(0, 1)$$

$$a) P[x > 10] = P[x' \geq 10,5] = P\left[z \geq \frac{10,5 - 9,6}{3,09}\right] = P[z \geq 0,29] = 1 - P[z < 0,29] = 1 - 0,6141 = 0,3859$$

b) Para una probabilidad del 95 %, $z_{\alpha/2} = 1,96$.

El intervalo característico será:

$$(9,6 - 1,96 \cdot 3,09; 9,6 + 1,96 \cdot 3,09) = (3,54; 15,66)$$

2 Si tenemos un dado correcto y lo lanzamos 50 veces:

a) ¿Qué probabilidad hay de que salgan más de diez unos?

b) ¿Cuál es la probabilidad de que salga «múltiplo de 3» al menos veinte veces?

a) Llamamos $x =$ «n.º de veces que sale el 1»; así, x es $B\left(50, \frac{1}{6}\right)$.

Como $np > 5$ y $nq > 5$, podemos aproximar mediante una normal de media $\mu = 50 \cdot \frac{1}{6} = 8,33$ y desviación típica $\sigma = \sqrt{50 \cdot \frac{1}{6} \cdot \frac{5}{6}} = 2,64$; es decir:

$$x \text{ es } B\left(50, \frac{1}{6}\right) \rightarrow x' \text{ es } N(8,33; 2,64) \rightarrow z \text{ es } N(0, 1)$$

$$P[x > 10] = P[x' \geq 10,5] = P\left[z \geq \frac{10,5 - 8,33}{2,64}\right] = P[z \geq 0,82] = 1 - P[z < 0,82] = 1 - 0,7939 = 0,2061$$

b) Llamamos $x =$ «n.º de veces que sale múltiplo de 3». La probabilidad de obtener un múltiplo de 3 en una tirada es $p = \frac{2}{6} = \frac{1}{3}$. Así, x es $B\left(50, \frac{1}{3}\right)$.

Como $np > 5$ y $nq > 5$, podemos aproximar mediante una normal de media $\mu = 50 \cdot \frac{1}{3} = 16,67$ y desviación típica $\sigma = \sqrt{50 \cdot \frac{1}{3} \cdot \frac{2}{3}} = 3,33$; es decir:

$$x \text{ es } B\left(50, \frac{1}{3}\right) \rightarrow x' \text{ es } N(16,67; 3,33) \rightarrow z \text{ es } N(0, 1)$$

$$P[x \geq 20] = P[x' \geq 19,5] = P\left[z \geq \frac{19,5 - 16,67}{3,33}\right] = P[z \geq 0,85] = 1 - P[z < 0,85] = 1 - 0,8023 = 0,1977$$

2 ► DISTRIBUCIÓN DE LAS PROPORCIONES MUESTRALES

Página 313

1 Como sabemos, en un dado correcto la proporción de veces que sale el 5 es $1/6 = 0,1\bar{6}$. Halla cada uno de los intervalos característicos correspondientes al 90 %, 95 % y 99 % para la «proporción de cincos», en tandas de 100 lanzamientos de un dado correcto.

Las proporciones de cincos en tandas de 100 lanzamientos siguen una distribución normal de media

$$p = \frac{1}{6} = 0,17 \text{ y desviación típica } \sqrt{\frac{pq}{n}} = \sqrt{\frac{(1/6) \cdot (5/6)}{100}} = 0,037; \text{ es decir, } pr \text{ es } N(0,17; 0,037).$$

Hallamos los intervalos característicos:

- Para el 90 %: $(0,17 \pm 1,645 \cdot 0,037) = (0,109; 0,231)$
- Para el 95 %: $(0,17 \pm 1,96 \cdot 0,037) = (0,097; 0,243)$
- Para el 99 %: $(0,17 \pm 2,575 \cdot 0,037) = (0,075; 0,265)$

3 ▶ INTERVALO DE CONFIANZA PARA UNA PROPORCIÓN O UNA PROBABILIDAD

Página 315

1 Se ha lanzado un dado 400 veces y se ha obtenido 72 veces el valor 4.

Estima el valor de la probabilidad $P[4]$ con un nivel de confianza del 90%.


Para un nivel de confianza del 90%, tenemos que $z_{\alpha/2} = 1,645$. La proporción de cuatros obtenidas en la muestra es:

$$pr = \frac{72}{400} = 0,18$$

El intervalo de confianza para estimar $P[4]$ será:

$$\left(0,18 - 1,645 \cdot \sqrt{\frac{0,18 \cdot 0,82}{400}}; 0,18 + 1,645 \cdot \sqrt{\frac{0,18 \cdot 0,82}{400}} \right) = (0,148; 0,212)$$

Es decir, con un nivel de confianza del 90%, la probabilidad de obtener 4 está entre 0,148 y 0,212.

2  [El ejercicio permite trabajar la creación y creatividad (dimensión personal) de esta clave].

¿Cuántas veces tendremos que lanzar un dado, que suponemos levemente incorrecto, para estimar la probabilidad de «SACAR 6» con un error menor que 0,002 y un nivel de confianza del 95%?

Para un nivel de confianza del 95%, tenemos que $z_{\alpha/2} = 1,96$. Como desconocemos el valor de pr , tomaremos $pr = \frac{1}{6} \approx 0,17$ (suponemos el dado levemente incorrecto).

El error máximo admisible es:

$$E = z_{\alpha/2} \cdot \sqrt{\frac{pr(1-pr)}{n}} \rightarrow 0,002 = 1,96 \cdot \sqrt{\frac{0,17 \cdot 0,83}{n}} \rightarrow n = 135\,512,44$$

Deberemos lanzarlo, al menos, 135 513 veces.

4 ► CONTRASTE DE HIPÓTESIS PARA UNA PROPORCIÓN

Página 316

- 1 a) En el ejemplo anterior, comprueba que si usamos un nivel de significación del 1 %, no podremos rechazar la hipótesis de que el dado es correcto.
- b) Lanzamos una moneda 100 veces y obtenemos 60 caras. ¿Podremos aceptar la hipótesis de que la moneda es correcta con un nivel de significación del 5 %?

a) Si $\alpha = 0,01 \rightarrow z_{\alpha/2} = 2,575$ y el intervalo característico correspondiente será:

$$(0,167 - 2,575 \cdot 0,037; 0,167 + 2,575 \cdot 0,037) = (0,072; 0,262)$$

$0,25 \in (0,072; 0,262)$, luego no podremos rechazar la hipótesis de que el dado es correcto.

b) HIPÓTESIS: La moneda es correcta. Por tanto, $P[\text{CARA}] = \frac{1}{2}$.

RESULTADO EMPÍRICO (a partir de la muestra): $pr(\text{cara}) = 0,6$.

Si la hipótesis fuera cierta, entonces las proporciones, pr , de «caras» en las muestras de tamaño 100 seguirían una distribución normal:

$$N\left(0,5; \sqrt{\frac{0,5 \cdot 0,5}{100}}\right) = N(0,5; 0,05)$$

Si $\alpha = 0,05 \rightarrow z_{\alpha/2} = 1,96$ y el intervalo característico correspondiente será:

$$(0,5 - 1,96 \cdot 0,05; 0,5 + 1,96 \cdot 0,05) = (0,402; 0,598)$$

$0,6 \notin (0,402; 0,598)$, luego no podremos aceptar la hipótesis de que la moneda es correcta.

EJERCICIOS Y PROBLEMAS RESUELTOS

Página 317

1. Distribución de las proporciones muestrales

Hazlo tú

- Halla la probabilidad de que el número de microcircuitos defectuosos en un paquete sea superior a 25.

Si hay más de 25 defectuosos, entonces la proporción de defectuosos es mayor que $\frac{25}{500} = 0,05$.

$$P[pr > 0,05] = P\left[z > \frac{0,05 - 0,04}{0,00876}\right] = P[z > 1,14] = 1 - P[z \leq 1,14] = 1 - 0,8729 = 0,1271$$

2. Estimación de una probabilidad

Hazlo tú

- Hemos fabricado, toscamente, un dado de madera. Lo lanzamos 400 veces y obtenemos 90 veces el 6. Estima la probabilidad de «SACAR 6» mediante intervalos con nivel de confianza:

a) del 90 %.

b) del 95 %.

c) del 99 %.

$$pr = \frac{90}{400} = 0,225$$

La desviación típica es: $s = \sqrt{\frac{0,225 \cdot 0,775}{400}} = 0,021$

a) $1 - \alpha = 0,90 \rightarrow z_{\alpha/2} = 1,645$

La cota de error es: $E = 1,645 \cdot 0,021 = 0,03455$

Con lo que el intervalo de confianza correspondiente a un nivel de confianza del 90 % queda:

$$(0,225 \pm 0,035) = (0,19; 0,26)$$

b) $1 - \alpha = 0,95 \rightarrow z_{\alpha/2} = 1,96$

La cota de error es: $E = 1,96 \cdot 0,021 = 0,041$

Con lo que el intervalo de confianza correspondiente a un nivel de confianza del 95 % queda:

$$(0,225 \pm 0,041) = (0,184; 0,266)$$

c) $1 - \alpha = 0,99 \rightarrow z_{\alpha/2} = 2,575$

La cota de error es: $E = 2,575 \cdot 0,021 = 0,054$

Con lo que el intervalo de confianza correspondiente a un nivel de confianza del 99 % queda:

$$(0,225 \pm 0,054) = (0,171; 0,279)$$

3. Tamaño de la muestra para estimar una proporción

Hazlo tú

- **Suponemos, en principio, que una moneda es correcta. ¿Cuántas veces habremos de lanzarla para estimar $P[C] = p$ con un error menor que 0,02 y con un nivel de confianza del 99%?**

$$1 - \alpha = 0,99 \rightarrow z_{\alpha/2} = 2,575$$

$$E = 2,575 \cdot \sqrt{\frac{0,5 \cdot 0,5}{n}} \rightarrow 0,02 = 2,575 \cdot \sqrt{\frac{0,5 \cdot 0,5}{n}} \rightarrow n = \frac{2,575^2 \cdot 0,25}{0,02^2} = 4\,144,1$$

Tendríamos que lanzar la moneda 4 145 veces para estimar la probabilidad con menos de dos centésimas de error y con un nivel de confianza del 99%.

4. Estimación de una proporción (resolución del problema inicial)

Hazlo tú

- **Estima el número de peces con un nivel de confianza del 80%.**

$$1 - \alpha = 0,8 \rightarrow z_{\alpha/2} = 1,28$$

$$E = 1,28 \cdot \sqrt{\frac{0,072 \cdot 0,928}{514}} = 0,015$$

Por tanto, el intervalo de confianza para p , al 80%, es: $(0,072 \pm 0,015) = (0,057; 0,087)$

$$0,057 = \frac{349}{N_1} \rightarrow N_1 \approx 6\,123$$

$$0,087 = \frac{349}{N_2} \rightarrow N_2 \approx 4\,011$$

Así, tenemos un nivel de confianza del 80% de que el número de peces del pantano esté en el intervalo $[4\,011, 6\,123]$.

EJERCICIOS Y PROBLEMAS GUIADOS

Página 319

1. Cálculo de probabilidades en una binomial mediante el paso a la normal

- En una distribución binomial

$$x \sim B(80; 0,11)$$

hallar $P[x > 4]$ pasando a una normal:

$$x' \sim N(\mu, \sigma)$$

a) tomando $x' \geq 4,5$.

b) tomando $x' \geq 4$.

$$\left. \begin{aligned} \mu &= np = 80 \cdot 0,11 = 8,8 \\ \sigma &= \sqrt{npq} = \sqrt{80 \cdot 0,11 \cdot 0,89} = 2,80 \end{aligned} \right\} \rightarrow x': N(8,8; 2,8)$$

$$a) P[x > 4] \approx P[x' \geq 4,5] = P\left[z \geq \frac{4,5 - 8,8}{2,8}\right] = P[z \geq -1,54] = 0,9382$$

$$b) P[x > 4] \approx P[x' \geq 4] = P\left[z \geq \frac{4 - 8,8}{2,8}\right] = P[z \geq -1,71] = 0,9564$$

2. Describir la distribución de las proporciones muestrales a partir de la p poblacional

- Sabemos que la proporción de personas Rh⁺ es de 0,11. ¿Cómo se distribuyen las pr en muestras de tamaño 80? Hallarla razonadamente.

En 80 individuos, el número, n , de ellos que son Rh⁺ se distribuye $B(80; 0,11)$ y, por tanto, n es $N(8,8; 2,8)$.

La proporción de Rh⁺ entre los 80 individuos $pr = \frac{n}{80}$ es $N\left(0,11; \sqrt{\frac{0,11 \cdot 0,89}{80}}\right) = N(0,11; 0,035)$.

3. Intervalo de confianza para p a partir de una muestra

- En una muestra de 80 personas hay 10 de ellas con Rh⁺.

Estimar p (proporción de Rh⁺ en la población) mediante un intervalo con un nivel de confianza del 95,5%.

$$pr = \frac{10}{80} = 0,125$$

$$s = \sqrt{\frac{0,125 \cdot 0,875}{80}} = 0,037$$

$$1 - \alpha = 0,955 \rightarrow \alpha = 1 - 0,955 = 0,045 \rightarrow \frac{\alpha}{2} = \frac{0,045}{2} = 0,0225$$

$$\rightarrow 1 - \frac{\alpha}{2} = 1 - 0,0225 = 0,9775 \rightarrow z_{\alpha/2} = 2,005$$

$$E = 2,005 \cdot 0,037 = 0,074$$

El intervalo es: $(0,125 \pm 0,074) = (0,051; 0,199)$

4. Número de individuos que debe tener una muestra

- Sabemos que la proporción de personas con Rh⁺ es un valor próximo a 0,1. Queremos estimar esta proporción en una etnia aún no estudiada.

¿Qué tamaño debe tener la muestra para que, con un nivel de confianza del 95 %, el error estadístico no sea superior a 0,002?

$$p = 0,1 \rightarrow s = \sqrt{\frac{0,1 \cdot 0,9}{n}}$$

Para un nivel de confianza del 95 % tenemos que $z_{\alpha/2} = 1,96$.

Por tanto, el error máximo es:

$$0,002 = 1,96 \cdot \sqrt{\frac{0,1 \cdot 0,9}{n}} \rightarrow n = \frac{1,96^2 \cdot 0,09}{0,002^2} = 86\,436$$

La muestra debe tener un tamaño de 86 436 personas.

5. Obtención del nivel de confianza de una estimación ya realizada

- En una muestra de 500 personas hemos obtenido una proporción $pr = 0,118$ de individuos con Rh⁺. Hacemos la estimación de que la proporción p de la población está en el intervalo (0,116; 0,120). ¿Con qué nivel de confianza hacemos esta estimación?

$$E = \frac{0,120 - 0,116}{2} = 0,002$$

$$pr = 0,118 \rightarrow s = \sqrt{\frac{0,118 \cdot 0,882}{500}} = 0,014$$

$$0,002 = z_{\alpha/2} \cdot 0,014 \rightarrow z_{\alpha/2} = \frac{0,002}{0,014} = 0,14$$

$$\frac{\alpha}{2} = P[z > 0,14] = 1 - 0,5557 = 0,4443 \rightarrow \alpha = 2 \cdot 0,4443 = 0,8886 \rightarrow 1 - \alpha = 1 - 0,8886 = 0,1114$$

El nivel de confianza es del 11 %.

EJERCICIOS Y PROBLEMAS PROPUESTOS

Página 320

Para practicar

Distribución de las proporciones muestrales

1 Averigua cómo se distribuyen las proporciones muestrales, p_r , para las poblaciones y las muestras que se describen a continuación:

| | a) | b) | c) | d) | e) | f) |
|--------------------------------------|-----|-----|-----|-----|------|------|
| PROPORCIÓN, p , EN LA POBLACIÓN | 0,5 | 0,6 | 0,8 | 0,1 | 0,05 | 0,15 |
| TAMAÑO, n , DE LA MUESTRA | 10 | 20 | 30 | 50 | 100 | 100 |

Recordemos que, si $np \geq 5$ y $nq \geq 5$, entonces, las proporciones muestrales siguen una distribución

$$N\left(p, \sqrt{\frac{pq}{n}}\right).$$

Aplicamos este resultado a cada uno de los casos propuestos. Observamos que en todos ellos se tiene que $np \geq 5$ y $nq \geq 5$.

- a) $N\left(0,5; \sqrt{\frac{0,5 \cdot 0,5}{10}}\right) = N(0,5; 0,158)$
 b) $N\left(0,6; \sqrt{\frac{0,6 \cdot 0,4}{20}}\right) = N(0,6; 0,110)$
 c) $N\left(0,8; \sqrt{\frac{0,8 \cdot 0,2}{30}}\right) = N(0,8; 0,073)$
 d) $N\left(0,1; \sqrt{\frac{0,1 \cdot 0,9}{50}}\right) = N(0,1; 0,042)$
 e) $N\left(0,05; \sqrt{\frac{0,05 \cdot 0,95}{100}}\right) = N(0,05; 0,0218)$
 f) $N\left(0,15; \sqrt{\frac{0,15 \cdot 0,85}{100}}\right) = N(0,15; 0,036)$

2 Halla los intervalos característicos para las proporciones muestrales del ejercicio anterior, correspondientes a las probabilidades que, en cada caso, se indican:

- a) 90% b) 95% c) 99% d) 95% e) 99% f) 80%

- a) $z_{\alpha/2} = 1,645$
Intervalo $(0,5 - 1,645 \cdot 0,158; 0,5 + 1,645 \cdot 0,158) = (0,24; 0,76)$
 b) $z_{\alpha/2} = 1,96$
Intervalo $(0,6 - 1,96 \cdot 0,110; 0,6 + 1,96 \cdot 0,110) = (0,38; 0,82)$
 c) $z_{\alpha/2} = 2,575$
Intervalo $(0,8 - 2,575 \cdot 0,073; 0,8 + 2,575 \cdot 0,073) = (0,61; 0,99)$
 d) $z_{\alpha/2} = 1,96$
Intervalo $(0,1 - 1,96 \cdot 0,042; 0,1 + 1,96 \cdot 0,042) = (0,018; 0,182)$
 e) $z_{\alpha/2} = 2,575$
Intervalo $(0,05 - 2,575 \cdot 0,0218; 0,05 + 2,575 \cdot 0,0218) = (-0,006; 0,106)$
 f) $z_{\alpha/2} = 1,28$
Intervalo $(0,15 - 1,28 \cdot 0,036; 0,15 + 1,28 \cdot 0,036) = (0,104; 0,196)$

3 El 42 % de los habitantes de un municipio es contrario a la gestión del alcalde y el resto son partidarios de este.

Si se toma una muestra de 64 individuos, ¿cuál es la probabilidad de que ganen los que se oponen al alcalde?

En muestras de 64, el número de personas que se oponen al alcalde, x , sigue una distribución binomial $B(64; 0,42)$.

Para ello, hemos de suponer que el municipio es suficientemente grande como para que, al ir tomando individuos para la muestra, la proporción no varíe sensiblemente. Es decir, cada individuo que extraigamos modifica la proporción. Pero si el número total es grande, esa variación es irrelevante.

Tenemos que calcular $P[x > 32]$. Como $np > 5$ y $nq > 5$, podemos aproximar mediante una normal de media:

$$\mu = n \cdot p = 64 \cdot 0,42 = 26,88$$

y desviación típica:

$$\sqrt{npq} = \sqrt{64 \cdot 0,42 \cdot 0,58} = 3,95$$

Así, si x es $B(64; 0,42) \rightarrow x'$ es $N(26,88; 3,95) \rightarrow z$ es $N(0, 1)$. Por tanto:

$$\begin{aligned} P[x > 32] &= P[x' \geq 32,5] = P\left[z \geq \frac{32,5 - 26,88}{3,95}\right] = P[z \geq 1,42] = \\ &= 1 - P[z < 1,42] = 1 - 0,9222 = 0,0778 \end{aligned}$$

4 La probabilidad de que un bebé sea varón es 0,515. Si han nacido 184 bebés, ¿cuál es la probabilidad de que haya 100 varones o más?

Halla el intervalo característico correspondiente al 95 % para la proporción de varones en muestras de 184 bebés.

- El número de varones entre 184 bebés, x , sigue una distribución binomial $B(184; 0,515)$. Tenemos que calcular $P[x \geq 100]$. Como $np > 5$ y $nq > 5$, podemos aproximar mediante una normal de media:

$$\mu = np = 184 \cdot 0,515 = 94,76$$

y desviación típica:

$$\sqrt{npq} = \sqrt{184 \cdot 0,515 \cdot 0,485} = 6,78$$

Así, si x es $B(184; 0,515) \rightarrow x'$ es $N(94,76; 6,78) \rightarrow z$ es $N(0, 1)$. Por tanto:

$$\begin{aligned} P[x \geq 100] &= P[x' \geq 99,5] = P\left[z \geq \frac{99,5 - 94,76}{6,78}\right] = P[z \geq 0,70] = \\ &= 1 - P[z < 0,70] = 1 - 0,7580 = 0,2420 \end{aligned}$$

- El intervalo característico para la proporción muestral es de la forma:

$$\left(p - z_{\alpha/2} \cdot \sqrt{\frac{pq}{n}}, p + z_{\alpha/2} \cdot \sqrt{\frac{pq}{n}}\right)$$

Para el 95 % $\rightarrow 1 - \alpha = 0,95 \rightarrow z_{\alpha/2} = 1,96$. Así, el intervalo será:

$$\left(0,515 - 1,96 \cdot \sqrt{\frac{0,515 \cdot 0,485}{184}}; 0,515 + 1,96 \cdot \sqrt{\frac{0,515 \cdot 0,485}{184}}\right) = (0,4428; 0,5872)$$

Intervalos de confianza

5 Se realizó una encuesta a 350 familias preguntando si poseían ordenador en casa, encontrándose que 75 de ellas lo poseían.

Estima la proporción real de las familias que disponen de ordenador con un nivel de confianza del 95 %.

La proporción de familias con ordenador en la muestra es $pr = \frac{75}{350} = \frac{3}{14}$.

Para el 95 % de confianza, $1 - \alpha = 0,95 \rightarrow z_{\alpha/2} = 1,96$.

El intervalo de confianza para p es:

$$\left(\frac{3}{14} - 1,96 \cdot \sqrt{\frac{(3/14)(1-(3/14))}{350}}; \frac{3}{14} + 1,96 \cdot \sqrt{\frac{(3/14)(1-(3/14))}{350}} \right) = (0,17; 0,26)$$

6 Se selecciona aleatoriamente una muestra de 600 personas en una ciudad y se les pregunta si consideran que el tráfico en la misma es aceptablemente fluido. Responden afirmativamente 250 personas.

¿Cuál es el intervalo de confianza de la proporción de ciudadanos de esa ciudad que consideran aceptable la fluidez del tráfico, con un nivel de confianza del 90 %?

La proporción muestral es:

$$pr = \frac{250}{600} = \frac{5}{12} \rightarrow 1 - pr = \frac{7}{12}$$

Para un nivel de confianza del 90 %, sabemos que $z_{\alpha/2} = 1,645$.

El intervalo de confianza para la proporción de ciudadanos que consideran aceptable la fluidez del tráfico es:

$$\left(pr - z_{\alpha/2} \cdot \sqrt{\frac{pr(1-pr)}{n}}, pr + z_{\alpha/2} \cdot \sqrt{\frac{pr(1-pr)}{n}} \right)$$

En este caso queda:

$$\left(\frac{5}{12} - 1,645 \cdot \sqrt{\frac{(5/12)(7/12)}{600}}; \frac{5}{12} + 1,645 \cdot \sqrt{\frac{(5/12)(7/12)}{600}} \right) = (0,3836; 0,4498)$$

Para resolver

7 En una determinada población, se sabe que el 14% de los individuos son daltónicos. Halla la probabilidad de que en un grupo de 500 personas de dicha población haya:

- a) 250 daltónicos b) ningún daltónico
c) algún daltónico d) más de 75 daltónicos.

El número de personas daltónicas, x , se puede aproximar por una distribución, X' :

$$N(500 \cdot 0,14; \sqrt{500 \cdot 0,14 \cdot 0,86}) \approx N(70; 7,76)$$

$$\begin{aligned} \text{a) } P[x = 250] &= P[250 - 0,5 < x' < 250 + 0,5] = P[249,5 < x' < 250,5] = P\left[\frac{249,5 - 70}{7,76} < x' < \frac{250,5 + 70}{7,76}\right] = \\ &= P[23,13 < z < 23,26] = 0 \end{aligned}$$

$$\text{b) } P[x = 0] = P[x' < 0,5]$$

De nuevo, al normalizar, es fácil ver que el valor será 0.

$$\text{c) } P[x > 0] = 1 - P[x = 0] = 1$$

$$\text{d) } P[x > 75] = P[x' > 75,5] = 1 - P[x' \leq 75,5] = 1 - P[z \leq 0,714] = 1 - 0,7612 = 0,2388$$

8 Sabemos que al lanzar al suelo 100 chinchetas, en el 95% de los casos, la proporción de ellas que quedan con la punta hacia arriba está en el intervalo (0,1216; 0,2784).

Di cuál es la probabilidad p de que una de esas chinchetas caiga con la punta hacia arriba y comprueba que la amplitud del intervalo dado es correcta.

- p es el centro del intervalo, es decir: $p = \frac{0,2784 + 0,1216}{2} = 0,2$

- Veamos que la amplitud del intervalo dado es correcta.

$$\text{Para el 95\% } \rightarrow 1 - \alpha = 0,95 \rightarrow z_{\alpha/2} = 1,96.$$

El intervalo característico es:

$$\left(p - z_{\alpha/2} \cdot \sqrt{\frac{pq}{n}}, p + z_{\alpha/2} \cdot \sqrt{\frac{pq}{n}} \right)$$

En este caso ($p = 0,2$; $q = 0,8$; $n = 100$; $z_{\alpha/2} = 1,96$), queda:

$$\left(0,2 - 1,96 \cdot \sqrt{\frac{0,2 \cdot 0,8}{100}}; 0,2 + 1,96 \cdot \sqrt{\frac{0,2 \cdot 0,8}{100}} \right) = (0,1216; 0,2784)$$

9 Se desea estimar la proporción de individuos que piensan votar a cierto partido político en una determinada ciudad. Para ello, se toma una muestra aleatoria de 300 individuos, resultando que 135 de ellos piensan votar a ese partido.

a) Calcula un intervalo de confianza al 95% para la proporción de individuos que piensan votar a ese partido.

b) Suponiendo que se mantiene la misma proporción muestral y el mismo nivel de confianza del apartado anterior, determina el tamaño mínimo de la muestra para estimar la proporción con un error inferior al 2%

$$\text{a) } pr = 135 / 300 = 0,45$$


Para una confianza del 95%, tenemos que $z_{\alpha/2} = 1,96$. El intervalo es:

$$\left(0,45 - 1,96 \cdot \sqrt{\frac{0,45 \cdot 0,55}{300}}; 0,45 + 1,96 \cdot \sqrt{\frac{0,45 \cdot 0,55}{300}} \right) = (0,39; 0,51)$$

b) Imponemos la igualdad:

$$E = 0,02 = 1,96 \cdot \sqrt{\frac{0,45 \cdot 0,55}{n}} \rightarrow n = 2376,99$$

El tamaño mínimo de la muestra deberá ser de 2377 individuos.

10  [La resolución del problema puede servir para trabajar esta técnica].

Se va a realizar un estudio de mercado para estimar la proporción de consumidores que conoce una determinada marca de yogures. Para ello, se va a tomar una muestra aleatoria simple de consumidores, se va a preguntar a cada uno si conoce la marca; y, a partir de los resultados, se construirá el intervalo de confianza correspondiente, a nivel de confianza del 90 %.

- a) Si queremos que la amplitud del intervalo de confianza sea menor que 0,08, ¿qué tamaño debe tener la muestra?
- b) Decidimos tomar una muestra de 175 consumidores, de los que 126 responden que conocen la marca. Calcula el intervalo de confianza al 90 % para la proporción de consumidores que conocen la marca.

a) $z_{\alpha/2} = 1,645$

Si la amplitud del intervalo debe ser menor que 0,08, el error debe ser menor que 0,04:

$$E = 1,645 \cdot \sqrt{\frac{pr \cdot (1 - pr)}{n}} \leq 0,04 \rightarrow 1691,27 \cdot pr \cdot (1 - pr) \leq n$$

Si nos ponemos en el peor de los casos, es decir, cuando $pr(1 - pr)$ es mayor, cosa que ocurre cuando $pr = 0,5$, la solución es $n \geq 423$.

b) $pr = 126 / 175 = 0,72$

$$1 - pr = 0,28$$

$$\left(0,72 - 1,645 \cdot \sqrt{\frac{0,72 \cdot 0,28}{175}}; 0,72 + 1,645 \cdot \sqrt{\frac{0,72 \cdot 0,28}{175}} \right) = (0,66; 0,78)$$

11 Se desea estimar la proporción, p , de individuos daltónicos de una población a través del porcentaje observado en una muestra aleatoria de individuos, de tamaño n .

- a) Si el porcentaje de individuos daltónicos en la muestra es igual al 30 %, calcula el valor de n para que, con un nivel de confianza del 95 %, el error cometido en la estimación sea inferior al 3,1 %.

- b) Si el tamaño de la muestra es de 64 individuos, y el porcentaje de individuos daltónicos en la muestra es del 35 %, determina, usando un nivel de confianza del 99 %, el correspondiente intervalo de confianza para la proporción de daltónicos de la población.

a) Para un nivel de confianza del 95 % $\rightarrow 1 - \alpha = 0,95 \rightarrow z_{\alpha/2} = 1,96$.

El error máximo admisible es: $E = z_{\alpha/2} \cdot \sqrt{\frac{pr(1 - pr)}{n}}$

Buscamos n para que $E = 0,031$.

$$1,96 \cdot \sqrt{\frac{0,3 \cdot 0,7}{n}} = 0,031 \rightarrow n = 839,48$$

La muestra ha de ser de 840 individuos.

- b) Para un nivel de significación del 1 %, tenemos que:

$$\alpha = 0,01 \rightarrow 1 - \alpha = 0,99 \rightarrow z_{\alpha/2} = 2,575$$

El intervalo de confianza para p será:

$$\left(0,35 - 2,575 \cdot \sqrt{\frac{0,35 \cdot 0,65}{64}}; 0,35 + 2,575 \cdot \sqrt{\frac{0,35 \cdot 0,65}{64}} \right) = (0,196; 0,504)$$

12 Un estudio publicado en una revista muestra que el 60 % de los 10 000 jóvenes de entre 16 y 25 años encuestados en diez países dijo estar extremadamente preocupado o muy preocupado por el cambio climático.

a) Calcula el intervalo de confianza al 99 % para la proporción de jóvenes de esos diez países que se sienten preocupados o muy preocupados por el cambio climático.

b) Si quisiéramos reducir el tamaño de la muestra, aunque aumentáramos el error a 0,02, ¿qué tamaño tendría?

a) Tomamos $pr = 0,6$.

Para el 99 %, $z_{\alpha/2} = 2,575$.

$$E = 2,575 \cdot \sqrt{\frac{0,6 \cdot 0,4}{10\,000}} = 0,0126$$

El intervalo es $(0,6 - 0,0126; 0,6 + 0,0126) = (0,5874; 0,6126)$

b) $0,02 = 2,575 \cdot \sqrt{\frac{0,6 \cdot 0,4}{n}} \rightarrow n = 2,575^2 \frac{0,6 \cdot 0,4}{0,02^2} = 3\,978,375$

Nos bastaría una muestra con 3 979 jóvenes.

13 En un aeropuerto, se toma una muestra de 100 días; en 25 de ellos hay saturación aérea. Con estos datos, se calculan dos intervalos de confianza para el parámetro *proporción de días con saturación aérea*: [0,122; 0,378] y [0,165; 0,335].

Justifica cuál es el intervalo de menor confianza.

La amplitud de los intervalos de confianza de la proporción es mayor cuanto mayor es $z_{\alpha/2}$, que a su vez es mayor cuanto mayor es el nivel de confianza.

Por tanto, como es menos amplio el intervalo [0,165; 0,335], es este el intervalo de menor nivel de confianza.

- 14** Un estudio electoral con una muestra de 400 electores, obtiene un intervalo para la proporción de votos para un partido de (0,23; 0,31). Averigua cuánto vale la proporción muestral, cuál es el nivel de confianza y cuál es el error máximo cometido en el intervalo anterior.

$$pr = \frac{0,23 + 0,31}{2} = 0,27$$

Por tanto:

$$(0,23; 0,31) = \left(0,27 - z_{\alpha/2} \cdot \sqrt{\frac{0,27 \cdot 0,73}{400}}; 0,27 + z_{\alpha/2} \cdot \sqrt{\frac{0,27 \cdot 0,73}{400}} \right)$$

Obtenemos que $z_{\alpha/2} = 1,80$, por tanto:

$$1 - \alpha = P[-1,80 < z < 1,80] = 2P[z < 1,80] - 1 = 2 \cdot 0,9641 - 1 = 0,9282$$

El nivel de confianza es del 92,82%:

$$\text{El error es } E = 1,80 \cdot \sqrt{\frac{0,27 \cdot 0,73}{400}} = 0,04.$$

Cuestiones teóricas

- 15** De una muestra de tamaño 400, se estima la proporción de individuos rubios en una ciudad. Se obtiene una cota de error de 0,0392 con un nivel de confianza del 95%.

a) ¿Podríamos, con la misma muestra, mejorar el nivel de confianza en la estimación? ¿A costa de qué?

b) ¿Sabrías hallar la proporción, pr , obtenida en la muestra?

a) Aumentando la cota de error mejoraría el nivel de confianza.

b) La cota de error es:

$$E = z_{\alpha/2} \cdot \sqrt{\frac{pr(1-pr)}{n}}$$

Como $E = 0,0392$; $n = 400$ y $1 - \alpha = 0,95 \rightarrow z_{\alpha/2} = 1,96$ tenemos que:

$$0,0392 = 1,96 \cdot \sqrt{\frac{pr(1-pr)}{400}} \rightarrow \frac{0,0392}{1,96} = \sqrt{\frac{pr(1-pr)}{400}} \rightarrow$$

$$\rightarrow 0,02 = \sqrt{\frac{pr(1-pr)}{400}} \rightarrow 0,0004 = \frac{pr(1-pr)}{400} \rightarrow$$

$$\rightarrow 0,16 = pr(1-pr) \rightarrow 0,16 = pr - pr^2 \rightarrow$$

$$\rightarrow pr^2 - pr + 0,16 = 0$$

$$pr = \frac{1 \pm \sqrt{1 - 0,64}}{2} = \frac{1 \pm \sqrt{0,36}}{2} = \frac{1 \pm 0,6}{2} \begin{cases} pr = 0,8 \\ pr = 0,2 \end{cases}$$

Podría ser $pr = 0,8$ o bien $pr = 0,2$. Con los datos que tenemos, no podemos decidir cuál de estos dos resultados es el válido.

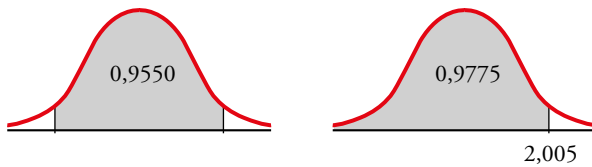
16 En una encuesta realizada a 800 personas elegidas al azar, 240 declaran su intención de votar al partido A.

- a) Estima, con un nivel de confianza del 95,5 %, entre qué valores se encuentra la intención de voto al partido A.
- b) Razona el efecto que tendría sobre el intervalo de confianza aumentar o disminuir el nivel de confianza.

La proporción muestral es:

$$pr = \frac{240}{800} = 0,3 \rightarrow 1 - pr = 0,7$$

a) Para un nivel de confianza del 95,5 %, hallamos $z_{\alpha/2}$:



$$1 - 0,9550 = 0,045; \frac{0,045}{2} = 0,0225$$

$$0,0225 + 0,9550 = 0,9775$$

$$P[z \leq z_{\alpha/2}] = 0,9775 \rightarrow z_{\alpha/2} = 2,005$$

El intervalo de confianza para estimar la proporción en la población es:

$$\left(pr - z_{\alpha/2} \cdot \sqrt{\frac{pr(1-pr)}{n}}, pr + z_{\alpha/2} \cdot \sqrt{\frac{pr(1-pr)}{n}} \right)$$

En este caso queda:

$$\left(0,3 - 2,005 \cdot \sqrt{\frac{0,3 \cdot 0,7}{800}}; 0,3 + 2,005 \cdot \sqrt{\frac{0,3 \cdot 0,7}{800}} \right) = (0,2675; 0,3325)$$

La proporción de votantes del partido A en la población se encuentra, con un nivel de confianza del 95,5 %, entre el 26,75 % y el 33,25 %.

- b) Si aumenta el nivel de confianza, mayor es la amplitud del intervalo; es decir, cuanto más seguros queramos estar de nuestra estimación, mayor será el error máximo admisible.
- Si disminuye el nivel de confianza, también lo hará la amplitud del intervalo.

Para profundizar

17 a) Un fabricante de medicamentos afirma que cierta medicina cura una enfermedad de la sangre en el 80% de los casos. Los inspectores de sanidad utilizan el medicamento en una muestra de 100 pacientes y deciden aceptar dicha afirmación si se curan 75 o más. Si lo que afirma el fabricante es realmente cierto, ¿cuál es la probabilidad de que los inspectores rechacen dicha afirmación?

b) Supongamos que en la muestra se curan 60 individuos. Di, con una confianza del 95%, cuál es el error máximo cometido al estimar que el porcentaje de efectividad del medicamento es del 60%.

a) Si lo que dice el fabricante es cierto, tenemos que $p = 0,8 \rightarrow 1 - p = 0,2$.

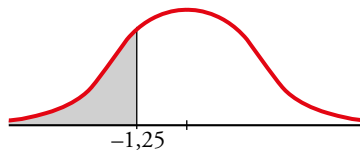
Considerando una muestra de tamaño $n = 100$, las proporciones muestrales, pr , siguen una distribución normal de media $p = 0,8$ y de desviación típica:

$$\sqrt{\frac{pq}{n}} = \sqrt{\frac{0,8 \cdot 0,2}{100}} = 0,04$$

es decir, pr es $N(0,8; 0,04)$.

La probabilidad de que los inspectores rechacen la afirmación es $P\left[pr < \frac{75}{100}\right]$.

Calculamos esta probabilidad:



$$\begin{aligned} P\left[pr < \frac{75}{100}\right] &= P[pr < 0,75] = P\left[z < \frac{0,75 - 0,8}{0,04}\right] = P[z < -1,25] = P[z > 1,25] = \\ &= 1 - P[z \leq 1,25] = 1 - 0,8944 = 0,1056 \end{aligned}$$

b) Si la proporción muestral es $pr = \frac{60}{100} = 0,6 \rightarrow 1 - pr = 0,4$.

Para $z_{\alpha/2} = 1,96$ (nivel de confianza del 95%), el error máximo será:

$$E = z_{\alpha/2} \cdot \sqrt{\frac{pr(1 - pr)}{n}} = 1,96 \cdot \sqrt{\frac{0,6 \cdot 0,4}{100}} \approx 0,096$$

El error máximo cometido es de un 9,6%, es decir, de 10 personas.

AUTOEVALUACIÓN

Página 321

1 En una población, la proporción de individuos que tienen una cierta característica es 0,32.

a) ¿Cómo se distribuyen las posibles proporciones, pr , de individuos con dicha característica en muestras de 200 individuos?

b) Halla el intervalo característico de pr correspondiente a un nivel de confianza del 95 %.

c) Calcula la probabilidad de que en una muestra la proporción sea menor que 0,3.

a) En la población, $p = 0,32$.

Las proporciones muestrales, pr , se distribuyen $N\left(p, \sqrt{\frac{pq}{n}}\right)$.

$$\sqrt{\frac{pq}{n}} = \sqrt{\frac{0,32 \cdot 0,68}{200}} = 0,033$$

Es decir, pr se distribuye $N(0,32; 0,033)$.

b) En una $N(0, 1)$, el intervalo característico correspondiente al 95 % es $(-1,96; 1,96)$.

$$0,32 - 1,96 \cdot 0,033 = 0,255$$

$$0,32 + 1,96 \cdot 0,033 = 0,647$$

El intervalo característico para pr (al 95 %) es $(0,255; 0,647)$.

$$c) P[pr < 0,3] = P\left[z < \frac{0,3 - 0,32}{0,033}\right] = P[z < -0,61] = 1 - \Phi(0,61) = 1 - 0,7291 = 0,2709$$

2 Se sabe que el 10 % de los habitantes de una determinada ciudad va regularmente al teatro. Se toma una muestra al azar de 100 habitantes de esta ciudad.

¿Cuál es la probabilidad de que, al menos, un 13 % de ellos vaya regularmente al teatro?

La distribución $x =$ «número de personas que van regularmente al teatro» es una $B(100; 0,1)$, donde $p = 0,1$ y $q = 1 - p = 0,9$.

Como $100 \cdot 0,1 > 5$ y $100 \cdot 0,9 > 5$, aproximamos con una distribución $x' \approx N(np, \sqrt{npq}) = N(10, 3)$, a la que aplicamos la corrección por continuidad:

$$\begin{aligned} P[x \geq 13] &= P[x' \geq 12,5] = P\left[z \geq \frac{12,5 - 10}{3}\right] = P[z \geq 0,83] = \\ &= 1 - \Phi(0,83) = 1 - 0,7967 = 0,2033 \end{aligned}$$

3 En una muestra de 60 estudiantes de una universidad, un tercio habla inglés.

a) Halla, con un nivel de confianza del 90 %, un intervalo para estimar la proporción de estudiantes que hablan inglés en esa universidad.

b) A la vista del resultado anterior, se va a repetir la experiencia para conseguir una cota de error de 0,01 con el mismo nivel de confianza. ¿Cuántos individuos deberá tener la muestra?

La proporción muestral es:

$$pr = \frac{1}{3} \rightarrow 1 - pr = \frac{2}{3}$$


Para un nivel de confianza del 90 %, sabemos que $z_{\alpha/2} = 1,645$.

a) El intervalo de confianza para estimar la proporción en la población es:

$$\left(pr - z_{\alpha/2} \cdot \sqrt{\frac{pr(1-pr)}{n}}, pr + z_{\alpha/2} \cdot \sqrt{\frac{pr(1-pr)}{n}} \right)$$

En este caso queda:

$$\left(\frac{1}{3} - 1,645 \cdot \sqrt{\frac{(1/3) \cdot (1/2)}{60}}; \frac{1}{3} + 1,645 \cdot \sqrt{\frac{(1/3) \cdot (1/2)}{60}} \right) = (0,2332; 0,4334)$$

4  **Meta 1.3.** [Tras visionar el vídeo el alumnado puede proponer medidas para mejorar la protección social para proteger a las personas más vulnerables].

Una encuesta realizada en cierto país sobre una muestra de 800 personas arroja el dato de que 300 son analfabetas. Para estimar la proporción de analfabetos del país, hemos obtenido el intervalo de confianza (0,3414; 0,4086). ¿Con qué nivel de confianza se ha hecho la estimación?

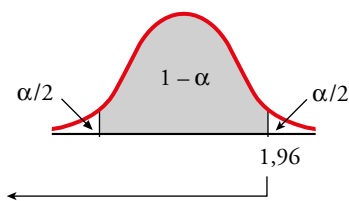
La proporción muestral es: $pr = \frac{300}{800} = \frac{3}{8} \rightarrow 1 - pr = \frac{5}{8}$

El error máximo admisible es la semiamplitud del intervalo de confianza; es decir:

$$E = \frac{0,4086 - 0,3414}{2} = 0,0336$$

Por tanto:

$$E = z_{\alpha/2} \cdot \sqrt{\frac{pr(1-pr)}{n}} \rightarrow 0,0336 = z_{\alpha/2} \cdot \sqrt{\frac{(3/8) \cdot (5/8)}{800}} \rightarrow z_{\alpha/2} = 1,96$$



$$P[z \leq 1,96] = 0,9750$$

$$\frac{\alpha}{2} = P[z > 1,96] = 1 - 0,9750 = 0,025$$

$$\alpha = 0,025 \cdot 2 = 0,05 \rightarrow 1 - \alpha = 0,95$$

El nivel de confianza es del 95 %.