MUESTREO

Actividad

Una fábrica produce ciertas piezas con una longitud media de 10 cm y una desviación típica de 1 cm.

a) ¿Cuál es la probabilidad de que la longitud media en una muestra de 50 piezas sea superior a 10,5 cm?
b) Si se toman 25 muestras de 50 piezas cada una, ¿en cuántas cabe esperar que la longitud media esté comprendida entre 9,8 cm y 10,3 cm?

Solución

a) Consideramos la variable aleatoria $\bar{X}$ que asigna a cada muestra de 50 piezas la media de sus longitudes.
El enunciado nos dice que $\mu = 10$, $\sigma = 1$ y $n = 50$.
Como el tamaño muestral es $n = 50 \geq 30$, podemos aproximar la distribución muestral de los medios por una normal:

$$\bar{X} \sim N\left(\mu, \frac{\sigma}{\sqrt{n}}\right) = N\left(10, \frac{1}{\sqrt{50}}\right) = N\left(10, \frac{1}{5\sqrt{2}}\right)$$

Por tanto, podemos encontrar la probabilidad pedida tipificando $\bar{X}$:

$$P(\bar{X} > 10,5) = P\left(\frac{\bar{X} - 10}{\frac{1}{5\sqrt{2}}} > \frac{10,5 - 10}{\frac{1}{5\sqrt{2}}}\right) =$$

$$= P(Z > 3,54) = 1 - P(Z \leq 3,54) =$$

$$= 1 - 0,9998 = 0,0002$$

La probabilidad pedida es 0,0002.

b) Consideramos la variable aleatoria $\bar{X}$ que asigna a cada muestra de 50 piezas la media de sus longitudes.
Sea $p = P(9,8 \leq \bar{X} \leq 10,3)$
Como la probabilidad es el límite de las frecuencias relativas cuando el número de veces, N, que se realiza un experimento aleatorio tiende a infinito, podemos tomar como frecuencia relativa del suceso A= «la longitud media de una muestra de 50 piezas está comprendida entre 9,8 y 10,3 cm», es decir, $(9,8 \leq \bar{X} \leq 10,3)$ el valor $f_A \simeq p$.
Por tanto, si n_A es el número de veces que se produce A al repetir el experimento aleatorio N = 25 veces (o sea, la frecuencia absoluta de A, que es lo que nos preguntan), se tiene:

$$\frac{n_A}{N} = f_A \simeq p \Rightarrow n_A \simeq N \cdot p = 25\,p$$

Para encontrar el valor de p, debemos conocer la distribución de $\bar{X}$.
Como $\mu=10$, $\sigma = 1$ y $n = 50 \geq 30$, podemos aproximar la distribución de $\bar{X}$ por una normal:

$$\bar{X} \sim N\left(\mu, \frac{\sigma}{\sqrt{n}}\right) = N\left(10, \frac{1}{\sqrt{50}}\right) = N\left(10, \frac{1}{5\sqrt{2}}\right)$$

Hallamos la probabilidad pedida tipificando $\bar{X}$

$$p = P\,(9{,}8 \le \overline{X} \le 10{,}3) =$$

$$= P\left(\frac{9{,}8 - 10}{\frac{1}{5\sqrt{2}}} \le \frac{\overline{X} - 10}{\frac{1}{5\sqrt{2}}} \le \frac{10{,}3 - 10}{\frac{1}{5\sqrt{2}}}\right) =$$

$$= P\,(-1{,}41 = 5\sqrt{2}\cdot(-0{,}2) \le Z \le 5\sqrt{2}\cdot 0{,}3 = 2{,}12) =$$

$$= P\,(Z \le 2{,}12) - P(Z \le -1{,}41) =$$

$$= P\,(Z \le 2{,}12) - (1 - P(Z \le 1{,}41)) =$$

$$= 0{,}9830 - 1 + 0{,}9207 = 0{,}9037$$

La probabilidad p es 0,9037, luego la frecuencia absoluta es $n_A \simeq N \cdot p = 25 \cdot 0{,}9037 = 22{,}59 \simeq 23$.
Así, cabe esperar que la longitud media esté comprendida entre 9,8 y 10,3 cm en 23 de las 25 muestras.

Actividad

¿Cuál es la probabilidad de que el peso medio en una muestra de 30 peras del ejemplo 1 sea mayor que 130 g?

Solución

El enunciado del ejemplo 1 nos dice que

$\mu = 125$, $\sigma = 20$.

Con una muestra de 30 peras tenemos que

$$\overline{X} \sim N\left(\mu, \frac{\sigma}{\sqrt{n}}\right) = N\left(125, \frac{20}{\sqrt{30}}\right) = N(125, 3{,}65)$$

Por tanto

$$P(\overline{X} > 130) = P\left(\frac{\overline{X} - 125}{3{,}65} > \frac{130 - 125}{3{,}65}\right)$$

$$= P(Z > 1{,}37) = 1 - P(Z \le 1{,}37)$$

$$= 1 - 0{,}9147 = 0{,}0853$$

La probabilidad que se nos pide es 0,0853.

Actividad

Una máquina fabrica bombillas que tienen una duración media de 700 horas y una desviación típica de 150 horas. ¿Cuál es la probabilidad de que la media de duración en una muestra de 100 bombillas sea menor o igual a 650 horas?

Solución

Según los datos del enunciado. Tenemos que $\mu = 700$, $\sigma = 150$ y $n = 100$.
Puesto que $n > 30$, podemos aproximar la distribución por una distribución normal

$$\overline{X} \sim N = \left(\mu, \frac{\sigma}{\sqrt{n}}\right) = N\left(700, \frac{150}{\sqrt{100}}\right) = N(700, 15)$$

La probabilidad que se nos pide es pues:

$$P(\overline{X} \leq 650) = P\left(\frac{\overline{X} - 700}{15} \leq \frac{650 - 700}{15}\right)$$
$$= P(Z \leq -3,33) = 1 - P(Z \leq 3,33)$$
$$= 1 - 0,9995 = 0,0005$$

La probabilidad que se nos pide es 0,0005.

Actividad

Al 75 % de los jóvenes de una ciudad les gusta el cine. Si seleccionamos 25 muestras de 100 jóvenes cada una, ¿en cuántas cabe esperar que el porcentaje de jóvenes cinéfilos esté comprendido entre el 70 % y el 80 %? ¿Y si las muestras fuesen de 1000 jóvenes?

— Compara ambos resultados y extrae conclusiones.

Solución

Consideramos la variable aleatoria $\hat{P}_n$ que asigna a cada muestra de n jóvenes la proporción a los que les gusta el cine.
Nos preguntan la frecuencia absoluta del suceso A = «el porcentaje de jóvenes cinéfilos está comprendido entre el 70 % y el 80 %» al tomar N = 25 muestras de n = 100 y n = 1 000 jóvenes, que se puede expresar $0,7 \leq \hat{P}_n \leq 0,8$.
Para calcularla, recordamos que la probabilidad es el límite de las frecuencias relativas, por lo que podemos aproximar la frecuencia relativa del suceso A por su probabilidad: $f_A \simeq p_A$.

De este modo, tenemos que:

$$\frac{n_A}{N} = f_A \simeq p_A \Rightarrow n_A \simeq N \cdot p_A = 25 \cdot p_A$$

Debemos hallar, pues, $p_A = P(0,7 \leq \hat{P}_n \leq 0,8)$ para n = 100 y n = 1000. Para ello, necesitamos conocer la distribución de $\hat{P}_n$.
El enunciado nos dice que la proporción poblacional de cinéfilos es p = 75 % = 0,75 y como n = 100 ≥ 30 y n = 1000 ≥ 30, entonces podemos aproximar la distribución de $\hat{P}_n$ por una normal:

$$\hat{P}_n \sim N\left(p, \sqrt{\frac{p \cdot (1-p)}{n}}\right) = N\left(0,75, \sqrt{\frac{0,75 \cdot 0,25}{n}}\right) =$$
$$= N\left(0,75, \frac{0,433}{\sqrt{n}}\right)$$

Por tanto, podemos tipificar la probabilidad pedida para $\hat{P}_n$:

$$p_A = P\ (0,7 \leq \hat{P}_n \leq 0,8) =$$
$$= P\left(\frac{0,7 - 0,75}{\frac{0,433}{\sqrt{n}}} \leq \frac{\hat{P}_n - 0,75}{\frac{0,433}{\sqrt{n}}} \leq \frac{0,8 - 0,75}{\frac{0,433}{\sqrt{n}}}\right) =$$
$$= P\ (-0,115\sqrt{n} \leq Z \leq 0,115\sqrt{n}) =$$
$$= 2 \cdot P\ (Z \leq 0,115\sqrt{n}) - 1$$

Si n = 100:

$$p_A = 2\,P\,(Z \leq 0{,}115 \cdot \sqrt{100}) - 1 =$$

$$= 2\,P\,(Z \leq 1{,}15) - 1 = 2 \cdot 0{,}8749 - 1 = 0{,}749\,8$$

Cabe esperar, pues, que si las muestras son de 100 individuos, el porcentaje de jóvenes cinéfilos esté comprendido entre el 70 % y el 80 % en $n_A \simeq N \cdot p_A = 25 \cdot 0{,}749\,8 = 18{,}745 \simeq 19$ de las 25 muestras.

Si n = 1 000:

$$p_A = 2\,P\,(Z \leq 0{,}115 \cdot \sqrt{1\,000}) - 1 =$$

$$= 2\,P\,(Z \leq 3{,}64) - 1 = 2 \cdot 0{,}999\,9 - 1 = 0{,}999\,8$$

Cabe esperar, pues, que si las muestras son de 1000 individuos, el porcentaje de jóvenes cinéfilos esté comprendido entre el 70 % y el 80 % en $n_A \simeq N \cdot p_A = 25 \cdot 0{,}999\,8 = 24{,}995 \simeq 25$ de las 25 muestras.
— Comprobamos que cuando el tamaño muestral es más grande es más probable que el valor que nos da el estadístico $\hat{P}$ esté comprendido entre 0,7 y 0,8, es decir, que esté más próximo al valor de la proporción poblacional p = 0,75.

Actividad

En el último año, el peso de los recién nacidos de dos maternidades *A* y *B* se ha distribuido normalmente (en la *A*, con media 3400 g y desviación típica 100 g; en la *B*, con media 3200 g y desviación típica 250 g). Halla la probabilidad de que, tomando muestras de 10 recién nacidos de cada maternidad, el peso medio de *A* supere al de *B* en más de 150 g.

Solución

Consideramos la variable aleatoria $\overline{X}_A - \overline{X}_B$ que asigna a cada par formado por una muestra de $n_A = 10$ recién nacidos en la maternidad A y una muestra de $n_B = 10$ recién nacidos en la maternidad B la diferencia de sus pesos medios.
El enunciado nos dice que el peso de los recién nacidos en la maternidad A sigue una normal N (μ_A, σ_A) = N (3400, 100), y el peso de los recién nacidos en la maternidad B sigue una normal N (μ_B, σ_B) = N (3200, 250), la variable $\overline{X}_A - \overline{X}_B$ sigue una distribución normal de parámetros:

$$\overline{X}_A - \overline{X}_B \sim$$

$$\sim N\left(\mu_A - \mu_B, \sqrt{\frac{\sigma_A^2}{n_A} + \frac{\sigma_B^2}{n_B}}\right) =$$

$$= N\left(3\,400 - 3\,200, \sqrt{\frac{100^2}{10} + \frac{250^2}{10}}\right) =$$

$$= N\,(200,\ 85{,}147)$$

Por tanto, la probabilidad de que el peso medio de los 10 recién nacidos en A supere al de B en más de 150 g, es decir, que $\overline{X}_A - \overline{X}_B > 150$, se obtiene tipificando $\overline{X}_A - \overline{X}_B$:

$$P(\overline{X}_A - \overline{X}_B > 150) =$$

$$= P\left(\frac{\overline{X}_A - \overline{X}_B - 200}{85{,}147} > \frac{150 - 200}{85{,}147}\right) =$$

$$= P\,(Z > -0{,}59) = P\,(Z \leq 0{,}59) = 0{,}722\,4$$

La probabilidad pedida es, pues, 0,7224.

Actividad

Considera la población formada por los números 1, 2, 3 y 4.

a) Calcula el parámetro poblacional μ.
b) Selecciona por muestreo aleatorio simple con reemplazamiento una muestra de tamaño 3 y, a partir de ella, estima μ usando $\bar{x}$ y usando *Me*. Compara tus estimaciones con el valor real y con las estimaciones obtenidas por tus compañeros.
c) Considera todas las posibles muestras de tamaño 3 seleccionadas como se indica en *b*, obtén los dos estimadores antes utilizados y halla su distribución de muestreo. Comprueba que los dos estimadores son insesgados y que el primero es más eficiente que el segundo.

Solución

a) La media poblacional es, por definición,

$$\mu = \frac{1+2+3+4}{4} = 2,5$$

b) Generamos números con la calculadora y nos quedamos con la primera cifra decimal, si ésta es un 1, 2, 3 ó 4, hasta tener 3 cifras:

$$\begin{array}{ccccccc} 0{,}299 & 0{,}896 & 0{,}127 & 0{,}741 & 0{,}078 & 0{,}944 & 0{,}420 \\ \downarrow & & \downarrow & & & & \downarrow \\ 2 & & 1 & & & & 4 \end{array}$$

Calculamos el valor de los estadísticos pedidos sobre esta muestra:

$$\bar{x} = \frac{1+2+4}{3} = 2,33$$

Me = 2, pues 2 es el valor central de la lista ordenada de cifras que forman la muestra: 1, 2, 4.
Las estimaciones obtenidas no se alejan en exceso del valor exacto μ = 2,5, pero aun así son demasiado groseras.

c) Escribimos en forma de tabla cada una de las posibles muestras que nos pueden salir utilizando muestreo aleatorio simple con reemplazamiento, de manera que importa el orden y hay repetición, junto alabr es uidlanst lx ¯ ysnda anestal eM

Muestra	$\bar{x}$	Me
1, 1, 1	1	1
1, 1, 2 1, 2, 1 2, 1, 1	1,33	
1 1 3 1 3 1 3 1 1	1,66	
1 1 4 1 4 1 4 1 1	2	

Muestra	$\bar{x}$	Me
1 2 2 2 1 2 2 2 1	1,66	2
1 2 3 1 3 2 2 1 3 2 3 1 3 1 2 3 2 1	2	

Muestra	$\bar{x}$	Me
1 2 4 1 4 2 2 1 4 2 4 1 4 1 2 4 2 1	2,33	2
1 3 3 3 1 3 3 3 1	2,33	3
1 3 4 1 4 3 3 1 4 3 4 1 4 1 3 4 3 1	2,66	

Muestra	$\bar{x}$	Me
1 4 4 4 1 4 4 4 1	3	4
2 2 2	2	2
2 2 3 2 3 2 3 2 2	2,33	
2 2 4 2 4 2 4 2 2	2,66	

Muestra	$\bar{x}$	Me
2 3 3 3 2 3 3 3 2	2,66	3
2 3 4 2 4 3 3 2 4 3 4 2 4 2 3 4 3 2	3	

Muestra	$\bar{x}$	Me
2 4 4 4 2 4 4 4 2	3,33	4
3 3 3	3	3
3 3 4 3 4 3 4 3 3	3,33	3
3 4 4 4 3 4 4 4 3	3,66	4
4 4 4	4	4

Actividad

Las puntuaciones obtenidas en ciertas oposiciones siguen una distribución *N* (μ, 25). Si en una muestra de 60 candidatos la media fue 85 puntos, halla intervalos de confianza para la puntuación media con niveles de confianza del 99 %, del 95 % y del 90 %. Fíjate en la amplitud de los intervalos obtenidos y extrae conclusiones.

Solución

Los datos del enunciado son:

$\bar{x} = 85$, $\sigma = 25$ y $n = 60$

Por tanto, el intervalo de confianza para la media poblacional con nivel de confianza $1 - \alpha$ es:

$$I_\alpha(\bar{x}) = \left[\bar{x} - z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}, \bar{x} + z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}\right] =$$

$$= \left[85 - z_{\frac{\alpha}{2}} \cdot \frac{25}{\sqrt{60}}, 85 + z_{\frac{\alpha}{2}} \cdot \frac{25}{\sqrt{60}}\right]$$

Así, para los niveles de confianza indicados:

- $1 - \alpha = 0,99 \Rightarrow z_{\frac{\alpha}{9}} = 2,58$, luego:

$$\left[85 - 2,58 \cdot \frac{25}{\sqrt{60}}, 85 + 2,58 \cdot \frac{25}{\sqrt{60}}\right] =$$
$$= [76,67, 93,33]$$

- $1 - \alpha = 0,95 \Rightarrow z_{\frac{\alpha}{2}} = 1,96$, luego:

$$\left[85 - 1,96 \cdot \frac{25}{\sqrt{60}}, 85 + 1,96 \cdot \frac{25}{\sqrt{60}}\right] =$$
$$= [78,67, 91,33]$$

- $1 - \alpha = 0,90 \Rightarrow z_{\frac{\alpha}{2}} = 1,65$, luego:

$$\left[85 - 1,65 \cdot \frac{25}{\sqrt{60}}, 85 + 1,65 \cdot \frac{25}{\sqrt{60}}\right] =$$
$$= [79,67, 90,33]$$

Observamos que cuanto menor es el nivel de confianza pedido, menor es el intervalo de confianza. Esto es lógico, pues cuanto más seguros queramos estar de que el parámetro pertenezca al intervalo, es decir, cuanto mayor sea el nivel de confianza, mayor debe ser dicho intervalo.

O sea:

$$1 - \alpha_1 < 1 - \alpha_2 \Rightarrow I_{\alpha_1}(\bar{x}) \subset I_{\alpha_2}(\bar{x})$$

Actividad

Para estudiar la proporción de estudiantes que practican determinado deporte, se toma una muestra de tamaño 300. El resultado obtenido es que 210 lo practican. Calcula el intervalo de confianza para la proporción *p* con nivel de confianza del 98 %.

Solución

Los datos del enunciado son:

$$n = 300,\ \hat{p} = \frac{210}{300} = 0{,}7 \ \text{y}\ 1-\alpha = 0{,}98$$

El valor crítico de nivel de confianza $1-\alpha = 0{,}98$ es $z_{\frac{\alpha}{2}}$ tal que:

$$P(Z \geq z_{\frac{\alpha}{2}}) = \frac{\alpha}{2} \Leftrightarrow P(Z \leq z_{\frac{\alpha}{2}}) = 1 - \frac{\alpha}{2} =$$

$$= 1 - \frac{0{,}02}{2} = 0{,}99$$

Consultando la tabla de la normal, obtenemos que $z_{\frac{\alpha}{2}} = 2{,}33$.

Así, el intervalo de confianza buscado es:

$$\left[0{,}7 - 2{,}33 \cdot \sqrt{\frac{0{,}7 \cdot (1-0{,}7)}{300}},\right.$$

$$\left. 0{,}7 + 2{,}33 \cdot \sqrt{\frac{0{,}7 \cdot (1-0{,}7)}{300}}\right] = [0{,}64,\ 0{,}76]$$

Actividad

Para saber si una moneda está o no trucada, se efectúan 700 lanzamientos y se obtienen 425 caras. Halla el intervalo de confianza para la proporción de caras obtenidas en sucesivos lanzamientos, con un nivel de confianza del 95 %. Según el resultado obtenido, razona si es de esperar o no que la moneda esté trucada.

Solución

Los datos del enunciado son:

$$n = 700,\ \hat{p} = \frac{425}{700} = 0{,}61 \ \text{y}\ 1-\alpha = 0{,}95$$

Calculamos el valor de $z_{\frac{\alpha}{2}}$ a partir de la tabla 1:

$$1-\alpha = 0{,}95 \Rightarrow z_{\frac{\alpha}{2}} = 1{,}96$$

Así, el intervalo de confianza buscado es:

$$\left[0,61 - 1,96 \cdot \sqrt{\frac{0,61 \cdot (1-0,61)}{700}},\right.$$

$$\left. 0,61 + 1,96 \cdot \sqrt{\frac{0,61 \cdot (1-0,61)}{700}}\right] = [0,57,\ 0,65]$$

La probabilidad de obtener cara si la moneda no está trucada es $\frac{1}{2}$. Pero:

$$\frac{1}{2} \notin [0,57,\ 0,65]$$

Así, podemos afirmar, con una probabilidad superior al 95 % de no equivocarnos, que la moneda está trucada.

Actividad

Una muestra de 50 bombillas de la marca *A* dio una vida media de 1500 h y una desviación típica de 100 h. Una muestra de 65 bombillas de la marca *B* dio una vida media de 1400 h y una desviación típica de 150 h. Halla el intervalo de confianza para la diferencia de medias de ambas marcas, con un nivel de significación del 6%. Si ambas marcas venden las bombillas al mismo precio, ¿cuáles conviene comprar? ¿Por qué?

Solución

Los datos del enunciado son:

$$\alpha = 0{,}06 \ ,\ \bar{x}_1 = 1\,500 \ ,\ \bar{x}_2 = 1\,400$$

$$\sigma_{n_1} = 100 \ ,\ \sigma_{n_2} = 150 \ ,\ n_1 = 50 \ \text{y}\ n_2 = 65$$

El valor crítico de nivel de confianza $1 - \alpha = 0{,}94$ es $z_{\frac{\alpha}{2}}$ tal que:

$$P(Z \geq z_{\frac{\alpha}{2}}) = \frac{\alpha}{2} \Leftrightarrow P(Z \leq z_{\frac{\alpha}{2}}) = 1 - \frac{\alpha}{2} = 0{,}97$$

y consultando la tabla de la normal tipificada, $z_{\frac{\alpha}{2}} = 1,89$.

Así, el intervalo de confianza para la diferencia de medias buscado es:

$$\left[0,7 - 2,33 \cdot \sqrt{\frac{0,7 \cdot (1-0,7)}{300}},\right.$$

$$\left. 0,7 + 2,33 \cdot \sqrt{\frac{0,7 \cdot (1-0,7)}{300}}\right] = [0,64,\ 0,76]$$

Esto significa que, con una probabilidad de error del 6%, la diferencia entre las duraciones medias de las bombillas de ambas marcas, μ_1 y μ_2 , está comprendida entre:

$55,45 \leq \mu_1 - \mu_2 \leq 144,55$

En particular, podemos afirmar que $\mu_1 \geq \mu_2 + 55$, lo que significa que cabe esperar que una bombilla de la marca A dure, por lo menos, 55 h más que una de la marca B. Puesto que las dos bombillas tienen el mismo precio, conviene comprar las de la marca A (pues duran más).

Actividad

Queremos saber el tiempo medio diario de estudio de los alumnos de Bachillerato españoles. En una muestra de 100 alumnos se obtuvo una desviación típica de 53 minutos. ¿Qué tamaño muestral hay que tomar para que el error máximo sea de 5 minutos, con una significación del 5%?

Solución

Los datos del enunciado son:

$$\alpha = 0{,}05 \ , \ E = 5 \ \text{y} \ \sigma_{n_1} = 53, \text{ donde } n_1 = 100$$

El valor crítico de nivel de confianza $1-\alpha = 0{,}95$ es $z_{\frac{\alpha}{2}} = 1{,}96$.

Como $n_1 = 100 \geq 30$, podemos aproximar la desviación típica poblacional por la desviación típica poblacional corregida:

$$\sigma \approx \hat{\sigma}_{n_1} = \sqrt{\frac{n_1}{n_1-1}} \cdot \sigma_{n_1} = \sqrt{\frac{100}{100-1}} \cdot 53 = 53{,}267$$

Así, para que el error máximo sea de 5 minutos:

$$5 = E = z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}} \simeq z_{\frac{\alpha}{2}} \cdot \frac{\hat{\sigma}_{n_1}}{\sqrt{n}} = 1{,}96 \cdot \frac{55{,}267}{\sqrt{n}} \Rightarrow$$

$$\Rightarrow \sqrt{n} = \frac{1{,}96 \cdot 53{,}267}{5} = 20{,}88 \Rightarrow n = 435{,}97$$

Para que el error máximo sea menor o igual que 5 minutos, debemos tomar n = 436.
Como hemos obtenido n = 436 ≥ 30, queda justificada la fórmula que hemos usado para el error máximo. Así, debemos tomar un tamaño muestral de 436 alumnos o más.

Actividad

En una muestra de 250 habitantes de una zona, 183 se manifestaron favorables a la apertura de un nuevo supermercado. ¿De qué tamaño deberá tomarse la muestra para que, con un nivel de confianza del 99 %, la proporción muestral y la poblacional no difieran en más de 0,02?

Solución

Los datos del enunciado son:

$$1-\alpha = 0{,}99 \ , \ \hat{p} = \frac{183}{250} = 0{,}732 \ \text{y} \ E = 0{,}02$$

El valor crítico de nivel de confianza $1-\alpha = 0{,}99$ es $z_{\frac{\alpha}{2}} = 2{,}58$.

Para que el error máximo de estimación (diferencia entre la proporción muestral y la poblacional, en valor absoluto) sea de 0,02:

$$0,02 = E = z_{\frac{\alpha}{2}} \cdot \sigma_{\hat{p}} =$$

$$= z_{\frac{\alpha}{2}} \cdot \sqrt{\frac{p(1-p)}{n}} \simeq z_{\frac{\alpha}{2}} \cdot \sqrt{\frac{\hat{p}\cdot(1-\hat{p})}{n}} =$$

$$= 2,58 \cdot \sqrt{\frac{0,732\cdot(1-0,732)}{n}} \Rightarrow$$

$$\Rightarrow \sqrt{n} = \frac{2,58\cdot\sqrt{0,732\cdot 0,268}}{0,02} = 57,14 \Rightarrow$$

$$\Rightarrow n = 3\,264,98$$

Como n debe ser entero, para que el error máximo sea menor o igual que 0,02 debemos tomar n = 3265.
Como n = 3265 ≥ 30 queda justificada la fórmula usada para el error máximo.
Deberíamos haber tomado una muestra de, al menos, n = 3265 habitantes.

Actividad

Una empresa desea averiguar el porcentaje de cerillas defectuosas de una partida. Para ello revisa 100 de las cerillas, y encuentra 8 defectuosas.

a) Con nivel de significación del 6 %, ¿qué error máximo puede cometerse generalizando el resultado obtenido a toda la partida?
b) ¿Cuántas cerillas más se deben revisar si queremos conseguir que el error máximo cometido sea de un 3%, con ese mismo nivel de significación?

Solución

a) Los datos del enunciado son:

$$n = 100\ ,\ \hat{p} = \frac{8}{100} = 0,08 \ \text{ y } \ \alpha = 0,06$$

Como estamos considerando proporciones y tenemos que n = 100 ≥ 30; el error máximo cometido viene dado por la fórmula:

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

El nivel crítico de nivel de confianza 1 – α = 0,94 es:

$$P(Z \geq z_{\frac{\alpha}{2}}) = \frac{\alpha}{2} \Rightarrow P(Z \leq z_{\frac{\alpha}{2}}) = 1 - \frac{\alpha}{2} = 0,97$$

Consultando las tablas de la normal tipificada, obtenemos $z_{\frac{\alpha}{2}} = 1,89$.

Finalmente, sustituimos el valor de n, $\hat{p}$ y $z_{\frac{\alpha}{2}}$ en la fórmula de error:

$$E = 1,89 \cdot \sqrt{\frac{0,08\cdot(1-0,08)}{100}} = 0,051 \Rightarrow$$

$$\Rightarrow E = 5,1\ \%$$

b) Podemos determinar el tamaño muestral n para que el error máximo cometido sea de un 3%.

Como el nivel de significación es el mismo, $z_{\frac{\alpha}{2}} = 1,89$, y como la única a... mos al valor de la proporción poblacional es $\hat{p} = 0,08$, tenemos que para... sea E = 3 %, el tamaño muestral n debe ser:

$$0,03 = E = z_{\frac{\alpha}{2}} \sqrt{\frac{\hat{p} \cdot (1 - \hat{p})}{n}} =$$

$$= 1,89 \cdot \sqrt{\frac{0,08 \cdot (1 - 0,08)}{n}} \Rightarrow$$

$$\Rightarrow n = \frac{1,89^2 \cdot 0,08 \cdot 0,92}{0,03^2} = 292,12$$

Puesto que n debe ser entero tomamos la aproximación entera por exceso, n = 293. Así, como ya hemos revisado 100 cerillas, debemos revisar 293 – 100 = 193 cerillas más para asegurar que el error máximo que hemos cometido es de un 3%.

Actividad

La nota de selectividad de los alumnos del centro *A* sigue una distribución *N* (6,5, 0,8), y la de los alumnos del centro *B*, *N* (6,1, 1,4). Seleccionando al azar 45 alumnos de *A* y 52 alumnos de *B*, ¿cuál es la probabilidad de que la nota media de los alumnos de *A* supere en más de 0,25 puntos a la nota media de los alumnos de *B*?

Solución

Sea $\bar{X}_A$ la variable aleatoria que asigna a cada muestra de $n_A = 45$ alumnos del centro A la media de sus notas, y $\bar{X}_B$ la que asigna a cada muestra de $n_B = 52$ alumnos del centro B la media de sus notas.
Como los alumnos del centro A se distribuyen según una normal N (6,5, 0,8) y los de los alumnos del centro B según una normal N (6,1, 1,4), la distribución de la variable aleatoria $\bar{X}_A - \bar{X}_B$ también es una normal, de parámetros:

$$\bar{X}_A - \bar{X}_B \sim N\left(6,5 - 6,1, \sqrt{\frac{0,8^2}{45} + \frac{1,4^2}{52}}\right) =$$

$$= N\ (0,4,\ 0,23)$$

Por tanto, como la nota media de los alumnos de A supera en más de 0,25 puntos a la nota media de los alumnos de B, sí y sólo si $\bar{X}_A - \bar{X}_B > 0,25$, podemos encontrar la probabilidad que nos piden tipificando:

$$P\ (\bar{X}_A - \bar{X}_B > 0,25) =$$

$$= P\left(\frac{\bar{X}_A - \bar{X}_B - 0,4}{0,23} > \frac{0,25 - 0,4}{0,23}\right) =$$

$$= P\ (Z > -0,65) = P\ (Z \leq 0,65) = 0,742\,2$$

Actividad

El número de horas diarias que duermen los estudiantes de Bachillerato de una cierta comunidad autónoma se distribuye según una ley normal de media desconocida y desviación típica 3. A partir de una muestra de tamaño 30, se ha obtenido una media muestral igual a 7 horas. Halla un intervalo de confianza al nivel del 96 % para la media de horas de sueño obtenida.

Solución

Los datos del enunciado son:

$\sigma = 3$, $n = 30$, $\bar{x} = 7$ y $1 - \alpha = 0{,}96$

El valor crítico de nivel de confianza 0,96 es:

$$P(Z > z_{\frac{\alpha}{2}}) = \frac{\alpha}{2} \Rightarrow P(Z \leq z_{\frac{\alpha}{2}}) = 1 - \frac{\alpha}{2} =$$

$$= 1 - \frac{0{,}04}{2} = 0{,}98$$

Consultando las tablas de la normal, obtenemos $z_{\frac{\alpha}{2}} = 2{,}06$.

Puesto que el número de horas que duermen los alumnos de Bachillerato, que es la característica de la población que se está estudiando, sigue una normal, el intervalo de confianza para la media con nivel de confianza 0,96 es:

$$\left[7 - 2{,}06 \cdot \frac{3}{\sqrt{30}}, 7 + 2{,}06 \cdot \frac{3}{\sqrt{30}}\right] =$$

$$= [5{,}872,\ 8{,}128]$$

Actividad

En las pruebas de selección de personal de una empresa, las puntuaciones obtenidas por los candidatos siguen una distribución *N* (μ, 35). Sabiendo que en una muestra de 50 candidatos se observó una media de 75 puntos, halla el intervalo de confianza para la puntuación media correspondiente a los niveles de confianza del 99 %, el 95 % y el 90 %. Compara la amplitud de los intervalos obtenidos y extrae conclusiones.

Solución

Los datos del enunciado son:

$\sigma = 35$, $n = 50$ y $\bar{x} = 75$

Puesto que la población (las puntuaciones obtenidas por los candidatos) tiene distribución normal, el intervalo de confianza para la media con nivel de confianza 1 – α es:

$$\left[75 - z_{\frac{\alpha}{2}} \cdot \frac{35}{\sqrt{50}}, 75 + z_{\frac{\alpha}{2}} \cdot \frac{35}{\sqrt{50}}\right] =$$

$$= [75 - 4{,}95 \cdot z_{\frac{\alpha}{2}},\ 75 + 4{,}95 \cdot z_{\frac{\alpha}{2}}]$$

Si $1 - \alpha = 0{,}99$, $z_{\frac{\alpha}{2}} = 2{,}58$, luego el intervalo de confianza es:

[75 – 4,95 · 2,58, 75 + 4,95 · 2,58] = [62,229, 87,771]

Si $1 - \alpha = 0{,}95$, $z_{\frac{\alpha}{2}} = 1{,}96$, luego el intervalo de confianza es:

[75 – 4,95 · 1,96, 75 + 4,95 · 1,96] = [65,298, 84,702]

Si $1 - \alpha = 0,9$, $z_{\frac{\alpha}{2}} = 1,65$, **luego el intervalo de confianza es:**

[75 – 4,95 · 1,65, 75 + 4,95 · 1,65] = [66,833, 83,168]

Observamos que $I_{0,01}(75) \supset I_{0,05}(75) \supset I_{0,10}(75)$. Ello es debido a que todos están centrados en $\bar{x} = 75$, pero cuanto menor sea el nivel de confianza, o sea la probabilidad con que queremos que el intervalo contenga el parámetro, más pequeños podemos tomar los intervalos.
Si lo pensamos al revés, está más claro: cuanto más grande sea un intervalo, mayor será la probabilidad de que contenga el parámetro (pues todos están centrados en 75), o sea el nivel de confianza.

Actividad

Para conocer las preferencias musicales de los habitantes de una población, se decide entrevistar al 5% de ellos usando muestreo aleatorio estratificado proporcional a la edad (niños-jóvenes-adultos). Si se ha calculado que esto supone entrevistar a 115 niños, 182 jóvenes y 398 adultos:

a) ¿Cuántos habitantes forman la población?
b) ¿Cuántos habitantes son niños? ¿Y jóvenes?

Solución

a) El número de individuos que se va a entrevistar es:

n = 115 + 182 + 398 = 695

Como se ha decidido entrevistar el 5% de la población, el número de individuos de la población será N tal que su 5% sea 695:

$$\frac{5}{100} \cdot N = 695 \Rightarrow \frac{695}{5} \cdot 100 = 13\,900$$

La población está formada, pues, por 13900 habitantes.

b) Puesto que la muestra es estratificada por edad, la proporción de individuos de cada grupo de edad en la población coincide con la proporción de dicho grupo en la muestra, lo cual nos permite calcular el número de niños n_1 y de jóvenes n_2 (y de adultos) de la población:

$$\frac{n_1}{13\,900} = \frac{115}{695} \Rightarrow n_1 = 2\,300$$

$$\frac{n_2}{13\,900} = \frac{182}{695} \Rightarrow n_2 = 3\,640$$

En la población hay, pues, 2300 niños y 3640 jóvenes.

Actividad

Para analizar el peso de unos botes de conserva, se toma una muestra de tamaño 32. Los pesos en kilogramos obtenidos son:

0,97	0,99	1,01	0,98	0,99	1,00	0,98	0,98
1,00	1,02	0,97	0,97	0,99	0,99	0,99	0,96
0,98	1,00	0,99	1,01	1,00	1,00	0,98	0,99
0,99	0,98	0,97	0,97	1,01	0,96	1,03	0,92

Halla el intervalo de confianza al nivel del 95% para el peso medio de los botes.

— Si el fabricante indica en los botes que el peso es de 1 kg, ¿es razonable pensar que está engañando a los clientes?

Solución

Los datos del enunciado son:

$n = 32$ y $1 - \alpha = 0,95$

Calculamos $\bar{x}$ a partir de la muestra: $\bar{x} = 0,987$

En este caso σ es desconocida, pero como $n = 32 \geq 30$ podemos considerar $\sigma \approx \hat{\sigma}_n$. A partir de los valores observados en la muestra se tiene que $\hat{\sigma}_n = 0,020$.

Finalmente:

$$1 - \alpha = 0,95 \Rightarrow z_{\frac{\alpha}{2}} = 1,96$$

Así, el intervalo de confianza buscado es:

$$\left[0,987 - 1,96\frac{0,020}{\sqrt{32}},\ 0,987 + 1,96\frac{0,020}{\sqrt{32}}\right] =$$

$$= [0,980,\ 0,994]$$

— Como el 95 % de los casos el peso indicado por los botes pertenece al intervalo de confianza y el peso marcado por el fabricante no es de dicho intervalo, $1 \notin [0,980,\ 0,994]$, no es razonable aceptar la afirmación del fabricante.

Actividad

La puntuación que obtienen los niños en cierto test psicológico sigue una distribución *N* (85, 15).

a) ¿Cuál es la probabilidad de que un niño elegido al azar obtenga más de 100 puntos?
b) ¿Cuál es la probabilidad de que la puntuación media en una muestra de 10 niños sea de más de 100 puntos?

Solución

a) Sea X la variable aleatoria que asigna a cada niño la puntuación que obtiene en el test.
El enunciado nos dice que $X \sim N(85, 15)$, por lo que podemos hallar la probabilidad de que un niño obtenga más de 100 puntos (o sea, de que $X > 100$) tipificando:

$$P(X > 100) = P\left(\frac{X - 85}{15} > \frac{100 - 85}{15}\right) =$$

$$= P(Z > 1) = 1 - P(Z \leq 1) =$$

$$= 1 - 0,8413 = 0,1587$$

b) Como la población es el conjunto de valores que toma una variable aleatoria normal, $X \sim N(85, 15)$, la variable aleatoria $\bar{X}$ que asigna a cada muestra de 10 niños la media de sus puntuaciones sigue una normal:

$$\bar{X} \sim N\left(85, \frac{15}{\sqrt{10}}\right)$$

Luego podemos hallar la probabilidad de que la puntuación media de los niños de la muestra sea mayor que 100 puntos (o sea, de que $\bar{X} > 100$) tipificando:

$$P(\bar{X} > 100) = P\left(\frac{\bar{X} - 85}{\frac{15}{\sqrt{10}}} > \frac{100 - 85}{\frac{15}{\sqrt{10}}}\right) =$$

$$= P\ (Z > \sqrt{10}) = 1 - P\ (Z \leq \sqrt{10} = 3{,}16) =$$

$$= 1 - 0{,}999\,2 = 0{,}000\,8$$

Actividad

Se desea comprobar la eficacia de dos tipos de somnífero en pacientes con insomnio. El somnífero *A* dio, en una muestra de 60 pacientes, una media de 7,15 h de sueño, con desviación típica de 0,65 h. El somnífero *B* dio, en una muestra de 80 pacientes, una media de 6,85 h de sueño, con desviación típica de 1,15 h. Halla el intervalo de confianza para la diferencia de medias, con un nivel de significación del 5%.

— ¿Es razonable aceptar que ambos somníferos son igual de eficaces? Razona tu respuesta.

Solución

Los datos del enunciado son:

$$n_1 = 60\ ,\ \bar{x}_1 = 7{,}15\ ,\ \sigma_{n_1} = 0{,}65\ ,\ n_2 = 80$$
$$\bar{x}_2 = 6{,}85\ ,\ \sigma_{n_2} = 1{,}15\ \text{ y }\ \alpha = 0{,}05$$

Calculamos

$z_{\frac{\alpha}{2}}$:

$$1 - \alpha = 0{,}95 \Rightarrow z_{\frac{\alpha}{2}} = 1{,}96$$

Como $n_1 = 60 \geq 30$ y $n_2 = 80 \geq 30$, podemos aproximar las desviaciones típicas muestrales por:

$$\sigma_1 \simeq \hat{\sigma}_{n_1} = \sqrt{\frac{n_1}{n_1 - 1}} \cdot \sigma_{n_1}$$

$$\sigma_2 \simeq \hat{\sigma}_{n_2} = \sqrt{\frac{n_2}{n_2 - 1}} \cdot \sigma_{n_2}$$

De este modo, el intervalo de confianza es:

$$\left[x_1 - x_2 - z_{\frac{\alpha}{2}} \cdot \sqrt{\frac{\frac{n_1}{n_1 - 1}\sigma_{n_1}^2}{n_1} + \frac{\frac{n_2}{n_2 - 1}\sigma_{n_2}^2}{n_2}}\right.,$$

$$\left.\bar{x}_1 - \bar{x}_2 + z_{\frac{\alpha}{2}} \cdot \sqrt{\frac{\frac{n_1}{n_1 - 1}\sigma_{n_1}^2}{n_1} + \frac{\frac{n_2}{n_2 - 1}\sigma_{n_2}^2}{n_2}}\right] =$$

$$= \left[\bar{x}_1 - \bar{x}_2 - z_{\frac{\alpha}{2}} \cdot \sqrt{\frac{\sigma_{n_1}^2}{n_1 - 1} + \frac{\sigma_{n_2}^2}{n_2 - 1}}\right.,$$

$$\left.x_1 - x_2 + z_{\frac{\alpha}{2}} \cdot \sqrt{\frac{\sigma_{n_1}}{n_1 - 1} + \frac{\sigma_{n_2}}{n_2 - 1}}\right]$$

Sustituyendo los datos:

$$\left[7{,}15 - 6{,}85 - 1{,}96 \cdot \sqrt{\frac{0{,}65^2}{60 - 1} + \frac{1{,}15^2}{80 - 1}}\right.,$$

$$\left.7{,}15 - 6{,}85 + 1{,}96 \cdot \sqrt{\frac{0{,}65^2}{60 - 1} + \frac{1{,}15^2}{80 - 1}}\right] =$$

$$= [-0{,}003,\ 0{,}603]$$

— Los dos somníferos son igual de eficaces si producen las mismas horas de sueño, es decir, si la diferencia de las medias de horas de sueño que producen es 0. Puesto que dicho valor pertenece al intervalo de confianza con nivel de significación 5%, $0 \in [-0{,}003,\ 0{,}603]$, debemos considerar razonable la suposición de que los dos son igual de eficaces.

Actividad

Un novelista desea conocer el porcentaje de habitantes de su ciudad que comprarán su próxima obra. Para ello, encarga a una empresa que entreviste a 100 habitantes. Sabiendo que 32 de ellos manifiestan querer adquirir la obra, averigua cuál es el error máximo cometido en la predicción y a cuántos habitantes más se ha de entrevistar para que, con nivel de confianza del 90 %, el error máximo sea del 5%.

Solución

Los datos del enunciado son:

$$n = 100\ ,\ \hat{p} = \frac{32}{100} = 032 \text{ y } 1 - \alpha = 0{,}9$$

Cálculo del nivel de confianza:

$$1 - \alpha = 0{,}9 \Rightarrow z_{\frac{\alpha}{2}} = 1{,}65\,.$$

Sustituyendo en la fórmula del error para la proporción se tiene:

$$E = 1{,}65 \cdot \sqrt{\frac{0{,}32 \cdot (1 - 0{,}32)}{100}} = 0{,}077 \Rightarrow E = 7{,}7\,\%$$

Para que el error fuese E = 5 %, el tamaño muestral debería ser:

$$0{,}05 = E = z_{\frac{\alpha}{2}} \cdot \sqrt{\frac{\hat{p} \cdot (1 - \hat{p})}{n}} =$$

$$= 1{,}65 \cdot \sqrt{\frac{0{,}32 \cdot (1 - 0{,}32)}{n}} \Rightarrow$$

$$\Rightarrow \sqrt{n} = \frac{1{,}65 \cdot \sqrt{0{,}32 \cdot 0{,}68}}{0{,}05} = 15{,}39 \Rightarrow n = 236{,}97$$

Para que el error sea menor o igual que 0,05, debemos entrevistar a n = 237 personas. Puesto que ya hemos entrevistado a 100 habitantes, debemos entrevistar a 237 – 100 = 137 habitantes más.

Actividad

El peso en gramos de ciertas magdalenas tiene distribución *N* (30, 5). Si esas magdalenas se envasan en bolsas de 16 unidades:

a) Halla la probabilidad de que, eligiendo una bolsa al azar, su contenido supere los 468 g.
b) El contenido del 67 % de las bolsas no supera cierto peso *x*. ¿De qué peso *x* se trata?

Solución

El peso total de las 16 magdalenas de una bolsa supera los 468 g si, y sólo si, el peso medio de las magdalenas de la bolsa supera los

$$\frac{468}{16} = 29{,}25 \text{ g.}$$

Abnabien, al ser la variable $\overline{X}$ que sigue la muestra de 16 magdalenas su peso medio sigue una distribución normal

$$N\left(30, \frac{5}{\sqrt{16}}\right) = N\,(30,\ 1{,}25),$$

pues el peso de las magdalenas sigue una normal N (30, 5).

Potr, podshalal ap bbilidad al espdaleménisde $\overline{X}$ ytpificad:

$$P\left(\sum_{i=1}^{16} x_i > 468\right) = P\left(\overline{x} = \frac{\sum_{i=1}^{16} x_i}{16} > \frac{468}{16} = 29{,}25\right) =$$

$$= P(\overline{X} > 29{,}25) = P\left(\frac{\overline{X} - 30}{1{,}25} > \frac{29{,}25 - 30}{1{,}25}\right) =$$

$$= P\,(Z > -0{,}6) = P\,(Z \leq 0{,}6) = 0{,}7257$$

b) Como la probabilidad es el límite de las frecuencias relativas, podemos reformular el enunciado: «Hallar x tal que la probabilidad de que el contenido de una bolsa (muestra de 16 magdalenas) no sea superior a x sea igual a 67 % = 0,67».
Razonandoágemetapt do, aepesesatcióneréinde $\overline{X}$ que sigue una N

(30, 1,25), y tipificaremos:

$$0,67 = P\left(\sum_{i=1}^{16} x_i \le x\right) = P\left(\bar{x} = \frac{\sum_{i=1}^{16} x_i}{16} \le \frac{x}{16}\right) =$$

$$= P\left(\bar{X} \le \frac{x}{16}\right) = P\left(\frac{\bar{X} - 30}{1,25} \le \frac{\frac{x}{16} - 30}{1,25}\right) =$$

$$= P\left(Z \le \frac{x - 480}{20}\right) \Rightarrow \frac{x - 480}{20} = 0,44 \Rightarrow$$

$$\Rightarrow x = 488,8 \text{ g}$$

Actividad

El salario medio de los trabajadores de un país es de 1 300 € mensuales, con una desviación típica de 150 €.
Los trabajadores de otro país cobran 1 200 € de media, con una desviación típica de 200 €. Se toman los datos del sueldo que cobran 100 trabajadores del primer país y 80 del segundo.
¿Cuál es la probabilidad de que el sueldo medio de los trabajadores de la primera muestra sea superior en más de 120 € al sueldo medio de los trabajadores de la muestra del segundo país?

Solución

Consideramos la variable aleatoria $\bar{X}_1 - \bar{X}_2$.
Los datos del enunciado son:

$\mu_1 = 1300$, $\mu_2 = 1200$, $\sigma_1 = 150$, $\sigma_2 = 200$, $n_1 = 100$, $n_2 = 80$

Como n_1 y n_2 son mayores de 30, se tiene que:

$$\bar{X}_1 - \bar{X}_2 \sim N\left(\mu_1 - \mu_2,\ \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}\right)$$

$$\bar{X}_1 - \bar{X}_2 \sim N\left(1300 - 1200,\ \sqrt{\frac{150^2}{100} + \frac{200^2}{80}}\right)$$

$$\bar{X}_1 - \bar{X}_2 \sim N\left(100,\ \sqrt{\frac{150^2}{100} + \frac{200^2}{80}}\right)$$

$$\bar{X}_1 - \bar{X}_2 \sim N(100,\ 26,926)$$

Si tipificamos la variable aleatoria:

$$Z = \frac{120 - 100}{26,926} = 0,743$$

$$P(\bar{X}_1 - \bar{X}_2 > 120) = P(Z > 0,743) = 1 - P(Z < 0,743)$$
$$P(\bar{X}_1 - \bar{X}_2 > 120) = 1 - 0,7704 = 0,2296$$

La probabilidad de que el sueldo medio de los trabajadores de la primera muestra sea superior en más de 120 € al sueldo medio de los trabajadores de la muestra del segundo país es del 23%.

Actividad

Sabiendo que la edad de los ancianos residentes de un geriátrico sigue una distribución *N* (74, 7), resuelve los siguientes apartados:

a) ¿Cuál es la probabilidad de que un anciano escogido al azar tenga más de 78 años?
b) ¿Qué porcentaje de ancianos tienen menos de 92 años?
c) ¿Qué porcentaje de ancianos tienen entre 70 y 92 años?
d) ¿Qué edad tiene una persona de esta residencia, sabiendo que el 35 % de los ancianos son más jóvenes que ella?

Solución

a) Consideramos la variable tipificada Z:

$$Z = \frac{78 - 74}{7} = 0,57$$

Por lo tanto:

$P(X > 78) = P(Z > 0,57) = 1 – P(Z \leq 0,57)$

Consultamos las tablas y obtenemos:

$P(X > 78) = 1 – P(Z \leq 0,57) = 1 – 0,7157 = 0,2843$

Por tanto hay una probabilidad del 28,43 % de que un anciano escogido al azar tenga más de 78 años.

b) Consideramos la variable tipificada Z:

$$Z = \frac{92 - 74}{7} = 2,57$$

Por lo tanto:

$P(X < 92) = P(Z < 2,57)$

Consultamos las tablas y obtenemos:

$P(X < 92) = P(Z < 2,57) = 0,9949$

El 99,49 % de los ancianos tiene menos de 92 años.

c) Consideramos las variables tipificadas Z_1 y Z_2:

$$Z_1 = \frac{70 - 74}{7} = -0,57 \qquad Z_2 = \frac{92 - 74}{7} = 2,57$$

Por lo tanto:

$P(70 < X < 92) = P(–0,57 \leq Z \leq 2,57) = P(Z \leq 2,57) – P(Z \leq 0,57)$

Consultamos las tablas y obtenemos:

$P(70 < X < 92) = P(Z \leq 2,57) – P(Z \leq 0,57) = 0,9949 – 0,2843 = 0,7106$

El 71,06 % de los ancianos tiene entre 70 y 92 años.

d) Hemos de hallar un número k tal que:

$P(X < k) = 0,35$

0,35 no aparece en la tabla. Buscaremos, pues, su simétrico. Miramos en la tabla el número que deja a su izquierda un área más cercana a 0,65 y éste es 0,39. Por tanto:

$$Z = \frac{k - 74}{7} = -0,39 \Rightarrow k = 71,27$$

Por tanto, una persona que tenga 71,27 años, (aproximadamente 71 años y tres meses), tiene un 35 % de los residentes más jóvenes que él.

Actividad

La velocidad a la que circulan los conductores por una autopista sigue una distribución *N* (μ, 20).
De un control de velocidad efectuado a 100 conductores al azar, ha resultado una velocidad media de 110 km/h. Determina el intervalo de confianza para la velocidad media, con un nivel de significación del 10 %.

Solución

Los datos del enunciado son:

$\alpha = 0,1$
$\bar{x} = 10$
$\sigma = 20$
$n = 100$

Por lo tanto, a partir de la tabla, para $1 - \alpha = 0,9$, tenemos que

$$z_{\frac{\alpha}{2}} = 1,65$$

Así pues, el intervalo de confianza buscado es:

$$\left[100 - 1,65 \cdot \frac{20}{\sqrt{100}}, 100 + 1,65 \cdot \frac{20}{\sqrt{100}}\right] =$$
$$= [96,7, 103,3]$$

Actividad

En un determinado concurso, los participantes van acumulando puntos a medida que van superando pruebas; y para poder pasar a la fase siguiente de pruebas, es necesario obtener 100 puntos o más. Sabiendo que la puntuación media obtenida por los concursantes en programas anteriores ha sido de 110 puntos con una desviación típica de 15 puntos, resuelve los apartados siguientes:

a) ¿Cuál es la probabilidad de que un concursante pase a la siguiente fase?
b) Si en la primera fase hay 10 concursantes, y sólo 3 pueden pasar a la segunda fase, ¿cuál es la puntuación que debe sacar un concursante para seguir participando?

Solución

a) La variable X: puntos obtenidos sigue una distribución normal N(110, 15).
La variable tipificada Z será:

$$Z = \frac{100 - 110}{15} = -0,67$$

Por lo tanto:

$P(X \geq 100) = P(Z \geq -0,67) = P(Z \leq 0,67)$

Consultamos las tablas y obtenemos:

$P(X > 100) = P(Z \leq 0,67) = 0,7486$

Por tanto hay una probabilidad del 74,86 % de que un concursante pase a la siguiente fase.

b) Solamente pueden pasar el

$$\left(\frac{3}{10} = 0,3\right) \; 30 \; \%$$

de los concursantes. Por lo tanto hemos de hallar un número k tal que:

$P(X < k) = 0,30$

0,30 no aparece en la tabla. Buscaremos, pues, su simétrico. Miramos en la tabla el número que deja a su izquierda un área más cercana a 0,7 y éste es 0,52. Por tanto:

$$Z = \frac{k - 110}{15} = 0,52 \Rightarrow k = 117,8$$

Se tendría que establecer una puntuación mínima de 118 puntos si se quiere que sólo pasen a la siguiente fase 3 de los 10 concursantes.

Actividad

Se está elaborando un estudio sobre el número de horas semanales que dedican los habitantes de una ciudad al desplazamiento desde su hogar hasta su lugar de trabajo.
Sabiendo que la desviación típica poblacional es de tres cuartos de hora, ¿de qué tamaño ha de ser la muestra poblacional para que, con una confianza del 99 %, la media muestral y la poblacional no difieran en más de 10 minutos?

Solución

Los datos que tenemos del enunciado, tomándolos en minutos son:

$1 - \alpha = 0,99$
$\sigma = 45$
$E = 10$

A partir de las tablas de la normal tipificada se tiene que:

$$1 - \alpha = 0,99 \Rightarrow z_{\frac{\alpha}{2}} = 2,58$$

Sustituyendo en la fórmula que nos da el error máximo se tiene:

$$E = z_{\frac{\alpha}{2}} \cdot \frac{\alpha}{\sqrt{n}} \quad \Rightarrow \quad 10 = 2,58 \cdot \frac{45}{\sqrt{n}} \quad \Rightarrow$$

$$\Rightarrow \quad \sqrt{n} = \frac{2,58 \cdot 45}{10} = 11,61 \quad \Rightarrow \quad n = 134,79$$

Por lo tanto, la muestra tendría que ser al menos de 135 personas.

Actividad

Los laboratorios fabricantes de un pesticida para combatir una plaga de un determinado cultivo, aseguran que es eficaz en un 90% de las aplicaciones.
Sabiendo que se ha preguntado a 200 agricultores que han probado el pesticida sobre los resultados obtenidos, ¿cuál es la probabilidad de que el pesticida haya funcionado en menos de 170 casos?

Solución

Consideramos la variable aleatoria $\hat{P}$ que asigna la proporción de casos en los que el pesticida ha funcionado.
Los datos del enunciado son:

p = 0,9 y n = 200

Como la muestra es suficientemente grande, podemos suponer que las proporciones siguen una distribución normal:

$$\hat{P} \sim N\left(p, \sqrt{\frac{p \cdot (1-p)}{n}}\right) = N\left(0{,}9, \sqrt{\frac{0{,}9 \cdot (1-0{,}9)}{200}}\right)$$

$$\hat{P} \sim N(0{,}9,\ 0{,}0212)$$

Se está preguntando por la probabilidad que el pesticida funcione en menos del 85 % de los casos

$$\frac{170}{200} = 0{,}85$$

Luego tipificando tenemos:

$$Z = \frac{0{,}85 - 0{,}9}{0{,}0212} = -2{,}358$$

$$P(\hat{P} < 0{,}85) = P(Z < -2{,}358) = P(Z > 2{,}358) =$$
$$= 1 - P(Z \leq 2{,}358) = 1 - 0{,}9909 = 0{,}0091$$

Así pues, la probabilidad de que haya funcionado en menos del 85 % de los casos es del 0,91 %.

Actividad

Sabiendo que si $n_1 \geq 30$ y $n_2 \geq 30$, la distribución de muestreo de la diferencia de proporciones es:

$$N\left(p_1 - p_2 \sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}\right)$$

indica la forma general del intervalo de confianza para la diferencia de proporciones.

— En una muestra de 150 jóvenes y 200 adultos, a 78 jóvenes y a 84 adultos les gusta cierto grupo musical. Determina el intervalo de confianza para la diferencia de proporciones con nivel de significación del 8%.

Solución

Sabemos que el intervalo de confianza para un parámetro λ con nivel de confianza 1 – α es:

$$I_\alpha\ (S\ (M)) = [S\ (M) - z_{\frac{\alpha}{2}} \cdot \sigma_S, S(M) + z_{\frac{\alpha}{2}} \cdot \sigma_S]$$

siendo M una muestra y S un estimador del parámetro que es insesgado y sigue una distribución normal N (λ, σ_S).

En este caso, nos interesa un intervalo de confianza para el parámetro diferencia de proporciones, $\lambda = p_1 - p_2$.

Como el enunciado nos dice que si $n_1 \geq 30$ y $n_2 \geq 30$ la diferencia de proporciones tiene distribución:

$$\hat{P}_1 - \hat{P}_2 \sim N\left(p_1 - p_2, \sqrt{\frac{p_1 \cdot (1-p_1)}{n_1} + \frac{p_2 \cdot (1-p_2)}{n_2}}\right)$$

pdemst oadesi adrS=P̂ $_1 - \hat{P}_2$ de la diferencia de proporciones para obtener un intervalo de confianza para dicha diferencia, siempre que $n_1 \geq 30$ y $n_2 \geq 30$ (pues es insesgado y sigue una distribución normal).

Enlpátic apsdonid,ersi n≥0, spued pxinppop̂

Así, la fórmula del intervalo de confianza que resulta es:

$$I_\alpha(\hat{p}_1 - \hat{p}_2) =$$

$$= \left[\hat{p}_1 - \hat{p}_2 - z_{\frac{\alpha}{2}} \cdot \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}},\right.$$

$$\left.\hat{p}_1 - \hat{p}_2 + z_{\frac{\alpha}{2}} \cdot \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}\right]$$

— Los datos del enunciado son:

$n_1 = 150$, $n_2 = 200$

$$\hat{p}_1 = \frac{78}{150} = 0,52 \text{ , } \hat{p}_2 = \frac{84}{200} = 0,42 \text{ y } \alpha = 0,08$$

El valor crítico del nivel de confianza $1 - \alpha = 1 - 0,08 = 0,92$ es:

$$P(Z > z_{\frac{\alpha}{2}}) = \frac{\alpha}{2} \Rightarrow P(Z \leq z_{\frac{\alpha}{2}}) = 1 - \frac{\alpha}{2} = 0,96$$

Consultando la tabla de la normal tipificada, obtenemos

$$z_{\frac{\alpha}{2}} = 1,76$$

Como $n_1 = 150 \geq 30$ y $n_2 = 200 \geq 30$, el intervalo de confianza para la diferencia de proporciones será:

$$\left[0,52 - 0,42 - 1,76 \cdot \sqrt{\frac{0,52 \cdot (1-0,52)}{150} + \frac{0,42 \cdot (1-0,42)}{200}},\right.$$

$$\left.0,52 - 0,42 + 1,76 \cdot \sqrt{\frac{0,52 \cdot (1-0,52)}{150} + \frac{0,42 \cdot (1-0,42)}{200}}\right]$$

$$= [0,006, 0,194]$$

Actividad

La media y la desviación típica de cierta población son $\mu = 30$ y $\sigma = 5$. Si $n = 64$, ¿cuál es el nivel de confianza del intervalo de confianza para la media [29,2, 30,8]?

Solución

Los datos del enunciado son:

$\mu = 30$, $\sigma = 5$ y $n = 64 \geq 30$

Por otro lado, sabemos que:

$$\bar{X} \sim N\left(\mu, \frac{\sigma}{\sqrt{n}}\right) = N\left(30, \frac{5}{8}\right)$$

Y, por tanto, el intervalo de confianza es de la forma:

$$\left[\bar{x} - z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}, \bar{x} + z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}\right] =$$

$$= \left[\bar{x} - z_{\frac{\alpha}{2}} \cdot \frac{5}{8}, \bar{x} + z_{\frac{\alpha}{2}} \cdot \frac{5}{8}\right]$$

Así, se cumple que:

$$\left.\begin{array}{l} \bar{x} - z_{\frac{\alpha}{2}} \cdot \frac{5}{8} = 29,2 \\ \bar{x} + z_{\frac{\alpha}{2}} \cdot \frac{5}{8} = 30,8 \end{array}\right\} \Rightarrow \bar{x} = 30, \; z_{\frac{\alpha}{2}} = 1,28$$

Por otro lado, si 1 – α es el nivel de confianza y Z ~ N (0, 1) se cumple que:

$$1 - \alpha = P\left(-z_{\frac{\alpha}{2}} \le Z \le z_{\frac{\alpha}{2}}\right) = 2P\left(z_{\frac{\alpha}{2}} \ge Z\right) - 1 =$$

$$= 2\,P(1,28 \ge Z) - 1 = 2 \cdot 0,8997 - 1 = 0,7994$$

Por tanto, el nivel de confianza para la media es del 79,94 % ≈ 80 %.

Actividad

Las calificaciones obtenidas por los alumnos de un curso universitario siguen una distribución normal de σ = 4. En un examen, la calificación media de una muestra de 100 alumnos ha sido 7. ¿Cuál es el intervalo de confianza para la media al nivel del 95%?

Solución

En primer lugar, tenemos que determinar el valor crítico $z_{\frac{\alpha}{2}}$. El valor crítico

$$Z \sim N(0, 1) \Rightarrow P\left(-z_{\frac{\alpha}{2}} \le Z \le z_{\frac{\alpha}{2}}\right) = 1 - \alpha = 0,95$$

Para obtener $z_{\frac{\alpha}{2}}$ a partir de una tabla de la normal tipificada calculamos:

$$0,95 = P\left(-z_{\frac{\alpha}{2}} \le Z \le z_{\frac{\alpha}{2}}\right) =$$

$$= P\left(-\infty < Z \le z_{\frac{\alpha}{2}}\right) - P\left(-\infty < Z < -z_{\frac{\alpha}{2}}\right) =$$

$$= P\left(-\infty < Z \le z_{\frac{\alpha}{2}}\right) - \left[1 - P\left(-\infty < Z \le z_{\frac{\alpha}{2}}\right)\right] =$$

$$= 2P\left(-\infty < Z \le z_{\frac{\alpha}{2}}\right) - 1$$

De donde:

$$0,95 = 2P\left(-\infty < Z \le z_{\frac{\alpha}{2}}\right) - 1 \Rightarrow$$

$$\Rightarrow P\left(-\infty < Z \le z_{\frac{\alpha}{2}}\right) = \frac{0,95 + 1}{2} = 0,975$$

Así, si consultamos la tabla de la normal tipificada se obtiene:

$$1 - \alpha = 0,95 \Rightarrow z_{\frac{\alpha}{2}} = 1,96$$

La distribución muestral de las medias sigue una distribución del tipo:

$$\bar{X} \sim N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$$

Calculamos

$$\frac{\sigma}{\sqrt{n}}:$$

$$\frac{\sigma}{\sqrt{n}} = \frac{4}{\sqrt{100}} = \frac{4}{10} = 0,4$$

La teoría nos dice que si $\bar{x}$ es la media de la muestra, el intervalo de confianza para la media de la población está dado por la fórmula:

$$\left[\bar{x} - z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}, \bar{x} + z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}\right]$$

Por tanto, los límites que fijan el intervalo de confianza serán:

$$\bar{x} - z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}} = 7 - 1,96 \cdot 0,4 = 6,2$$

$$\bar{x} + z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}} = 7 + 1,96 \cdot 0,4 = 7,8$$

El intervalo de confianza será [6,2, 7,8].

Actividad

Sabiendo que una determinada población sigue una distribución normal de desviación típica σ = 3, determina el intervalo de confianza, al nivel del 90 %, de una muestra de 200 datos con una media de 15.

Solución

En primer lugar, tenemos que determinar el valor crítico $z_{\frac{\alpha}{2}}$. El valor crítico $z_{\frac{\alpha}{2}}$

de confianza del 90 % es un número que con la normal tipificada $Z \sim N(1, 0)$

$$P\left(-z_{\frac{\alpha}{2}} \leq Z \leq z_{\frac{\alpha}{2}}\right) = 1 - \alpha = 0,9$$

Para obtener $z_{\frac{\alpha}{2}}$ a partir de una tabla de la normal tipificada hacemos:

$$\begin{aligned} 0,9 &= P\left(-z_{\frac{\alpha}{2}} \leq Z \leq z_{\frac{\alpha}{2}}\right) = \\ &= P\left(-\infty < Z \leq z_{\frac{\alpha}{2}}\right) - P\left(-\infty < Z < -z_{\frac{\alpha}{2}}\right) = \\ &= P\left(-\infty < Z \leq z_{\frac{\alpha}{2}}\right) - \left[1 - P\left(-\infty < Z \leq -z_{\frac{\alpha}{2}}\right)\right] = \\ &= 2P\left(-\infty < Z \leq z_{\frac{\alpha}{2}}\right) - 1 \end{aligned}$$

De donde:

$$0,9 = 2P\left(-\infty < Z \leq z_{\frac{\alpha}{2}}\right) - 1 \Rightarrow$$

$$\Rightarrow P\left(-\infty < Z \leq z_{\frac{\alpha}{2}}\right) = \frac{0,9+1}{2} = 0,95$$

En la tabla de la normal reducida encontramos que la probabilidad 0,95 corresponde a:

$$z_{\frac{\alpha}{2}} = 1,65$$

La distribución muestral de medias sigue una distribución del tipo:

$$\bar{X} \sim N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$$

Calculamos el valor de

$\frac{\sigma}{\sqrt{n}}$:

$$\frac{\sigma}{\sqrt{n}} = \frac{3}{\sqrt{200}} = \frac{3}{10\sqrt{2}} = 0,212$$

Latránsdcqesx̄ sl ardiadlanst dj nitdad cnfizpnalnda dea
población está dado por la fórmula:

$$\left[\bar{x} - z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}, \bar{x} + z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}\right]$$

Por tanto, los límites que fijan el intervalo de confianza serán:

$$\bar{x} - z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}} = 15 - 1,65 \cdot 0,212 = 14,65$$

$$\bar{x} + z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}} = 15 + 1,65 \cdot 0,212 = 15,35$$

Así, el intervalo de confianza es [14,65, 15,35].

todos