11 Inferencia estadística

Actividades propuestas

Muestreo y técnicas de encuestas

1. Diseña una técnica de muestreo para estimar el porcentaje de conductores de una ciudad que utilizan el móvil.

Se podría utilizar un muestreo estratificado por edades, sexo y tipo de vehículo. Posteriormente, un muestreo aleatorio simple dentro de cada estrato.

2. ¿Qué tipo de muestreo te parece más adecuado para conocer la proporción de extranjeros en un municipio?

Parece razonable suponer que no hay la misma proporción de extranjeros por edades. Por tanto, se efectuaría un muestreo estratificado por franjas de edades y, a continuación, un muestreo aleatorio simple dentro de cada estrato teniendo en cuenta el tamaño de la población.

3. Diseña un muestreo estratificado y por conglomerados para estimar la proporción de usuarios españoles del transporte público.

- Muestreo estratificado por edades y nivel de ingresos. Después, muestreo aleatorio simple por estratos.

- Muestreo por conglomerados seleccionando comunidades autónomas, municipios y zonas. Después, muestreo aleatorio simple en cada zona.

4. Investiga qué técnica de recogida de datos utilizan los barómetros del CIS.

Respuesta abierta.

Teoría central del límite

5. Chanclas. El número de kilómetros que puede recorrer una joven con unas chanclas, antes de que queden inservibles, es una variable aleatoria normal de media 20 km y una desviación típica de 4 km. ¿Cuál es el menor número de pares de chanclas que debe llevar en la mochila, para tener una garantía de, al menos, el 91 % de que no tendrá que caminar descalza si hace un recorrido de 300 km?

Sean x_i = «número de kilómetros que aguanta la joven en i pares de chanclas», con i = 1, 2… n. Y sea n el número de pares de chanclas. Entonces, la variable aleatoria suma S:

$$S = x_1 + x_2 + \cdots + x_n = \sum_{i=1}^{n} x_i$$

Como $x_i \sim N(20,4)$, entonces:

$$S \sim N(\mu_S = 20n, \sigma_S = 4\sqrt{n})$$

La probabilidad de tener, al mes, un 91 % de garantía de recorrer 300 km es:

$$p(s \geq 300) = 0{,}91 \rightarrow p(s < 300) = 0{,}09$$

$z = \frac{s-20n}{4\sqrt{n}} \sim N(0,1)$,

Luego $p\left(z < \frac{300-20n}{4\sqrt{n}}\right) = 0{,}09 \rightarrow \frac{300-20n}{4\sqrt{n}} = -1{,}34 \rightarrow n = 17$

Necesita, pues, al menos 17 pares de chanclas para cubrir sus expectativas.

6. Válvula de una máquina. Una máquina necesita para su funcionamiento una válvula de la que dispone en existencias de 300 unidades. La duración de las válvulas es de media 5 minutos y una desviación típica de 1 minuto. Si se avería una válvula es reemplazada automáticamente por otra. ¿Qué probabilidad tiene la máquina de funcionar durante, al menos, 25 horas?

Sea x_i = «duración, en minutos, de la válvula i-ésima» con i = 1, 2, … n

Como $x_i \sim N(5,1)$, entonces:

$$S = x_1 + x_2 + \cdots + x_{300} = \sum_{i=1}^{300} x_i$$

$$\mu_S = 300 \cdot S = 1500$$

$$\sigma_S = 1 \cdot \sqrt{300} = 10 \cdot \sqrt{3}$$

$$S \sim N(1500{,}10\sqrt{3}$$

$$z = \frac{s - 1500}{10\sqrt{3}} \sim N(0,1)$$

La probabilidad de que la máquina funcione durante, al menos, 25 horas = 1500 minutos, es:

$$p\left(S \geq 100) = p(z < \frac{1500 - 1500}{10\sqrt{3}}\right) = p(z \geq 0) = 0{,}5\ o\ 50\ \%$$

Distribución muestral de la media

7. **Una variable aleatoria sigue una distribución normal de media μ = 70 y una desviación típica = 9. Determina la probabilidad de que, en una muestra aleatoria simple de tamaño n = 36, tenga una media:**

a) Mayor de 68.

b) Entre 69 y 75.

Población $[N(\mu,\sigma)]$: $\begin{cases}\mu = 70\\ \sigma = 9\end{cases}$; muestra: n = 36

$$\mu_{\bar{x}} = \mu = 70;\ \sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{9}{6} = 1{,}5$$

$$\bar{x} \sim N(70;\ 1{,}5), o\ bien, z = \frac{\bar{x} - 70}{1{,}5} \sim N(0; 1)$$

a) $p(\bar{x} > 68) = p\left(z > \frac{68-70}{1{,}5}\right) = p(z > -1{,}33) = p(z < 1{,}33) = 0{,}9082$

b) $p(69 < \bar{x} < 75) = p\left(\frac{69-70}{1{,}5} < z < \frac{75-70}{1{,}5}\right) = p(-0{,}67 < z < 3{,}33) = 0{,}7486$

8. **La distribución de la media muestral para una muestra de tamaño 100 tiene una desviación típica igual a 2. ¿Cuál es la desviación típica de la población de partida?**

Población $[N(\mu,\sigma)]$: $\begin{cases}\mu = ?\\ \sigma = ?\end{cases}$; muestra: n = 100

Como $\sigma_{\bar{x}} = 2 \rightarrow \sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} \rightarrow 2 = \frac{\sigma}{10} \rightarrow \sigma = 2$

La desviación típica de la población de partida es 2.

9. Horas extra. **En una fábrica de automóviles el número de horas extra realizadas al año por los trabajadores fue de media 60 horas y una desviación típica de 12 horas. En una muestra aleatoria de 49 trabajadores, ¿cuál es la probabilidad de que el número medio de horas extra realizadas al año sea superior a 65? ¿Y entre 55 y 65 horas?**

Población $[N(\mu,\sigma)]$: $\begin{cases}\mu = 60\ horas\\ \sigma = 14\ horas\end{cases}$; muestra: n = 49 empleados

La desviación de la media muestral es:

$$\mu_{\bar{x}} = \mu = 60;\ \sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{14}{7} = 2$$

$$\bar{x} \sim N(60, 2), o\ bien, z = \frac{\bar{x} - 60}{2} \sim N(0; 1)$$

$$p(\bar{x} > 65) = p\left(z > \frac{65 - 60}{2}\right) = p(z > 2{,}5) = 1 - p(z \leq 2{,}5) = 1 - 0{,}9938 = 0{,}0062\ o\ 0{,}62\ \%$$

La probabilidad de que el número de horas extra realizadas al año sea superior a 65 horas es del 0,62 %.

$$p(55 < \bar{x} < 65) = p(-2{,}5 < z < 2{,}5) = 2p(z < 2{,}5) - 1 = 0{,}9876\ o\ 98{,}76\ \%$$

10. *Call center.* **Las llamadas recibidas en un *call center* tienen una duración media de 3,5 minutos y una desviación típica de 1,4 minutos. Un día determinado se elige una muestra aleatoria de 49 llamadas. ¿Cuál es la probabilidad de que la duración media oscile entre 3 y 4 minutos?**

Población $[N(\mu, \sigma)]$: $\begin{cases} \mu = 3{,}5\ minutos \\ \sigma = 1{,}4\ minutos \end{cases}$; muestra: n = 49 llamadas

La desviación de la media muestral es:

$$\mu_{\bar{x}} = \mu = 3{,}5;\ \sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{1{,}4}{7} = 0{,}2$$

$$\bar{x} \sim N(3{,}5; 0{,}2), o\ bien, z = \frac{\bar{x} - 3{,}5}{0{,}2} \sim N(0; 1)$$

$$p(3 < \bar{x} < 4) = p\left(\frac{3 - 3{,}5}{0{,}2} < z < \frac{4 - 3{,}5}{0{,}2}\right) = p(-2{,}5 < z < 2{,}5) = 2p(z < 2{,}5) - 1$$
$$= 0{,}9876\ o\ 98{,}76\ \%$$

Distribución muestral de una proporción

11. **Estudiantes aprobados. Se conoce, por cursos pasados, que la proporción de estudiantes que aprueban Matemáticas Aplicadas a las Ciencias Sociales II es del 84 %. En una muestra aleatoria de los expedientes de 40 estudiantes que cursan esa asignatura, ¿cuál es la probabilidad de que la proporción muestral de estudiantes que aprobaron esa asignatura superara el 85 %?**

Población $[N(\mu, \sigma)]$: $\begin{cases} \mu = 0{,}84 \\ \sigma = 0{,}16 \end{cases}$; muestra: n = 40

$$\mu_{\hat{p}} = p = 0{,}84;\ \sigma_{\hat{p}} = \sqrt{\frac{p \cdot q}{n}} = \sqrt{\frac{0{,}84 \cdot 0{,}16}{40}}$$

$$\hat{p} \sim N(0{,}84; 0{,}058), o\ bien, z = \frac{\hat{p} - 0{,}84}{0{,}058} \sim N(0; 1)$$

$$p(\hat{p} > 0{,}85) = p\left(z > \frac{0{,}85 - 0{,}84}{0{,}058}\right) = p(z > 0{,}17) = 1 - p(z \leq 0{,}17) = 1 - 0{,}0675$$
$$= 0{,}9325\ o\ 93{,}25\ \%$$

La probabilidad de que la proporción muestral de estudiantes que aprobaron esta asignatura superara el 85 % es del 93,25 %.

12. Comercios de barrio. Según los datos que maneja la asociación de empresarios de una ciudad, el 20 % de las personas realiza sus compras de alimentación de forma habitual en pequeños comercios o galerías comerciales de barrio. En una muestra elegida al azar de 50 personas, calcula la probabilidad de que la proporción de personas que realicen sus compras en pequeños comercios o galerías comerciales de barrio:

a) sea superior al 25 %

b) sea inferior al 22 %.

c) esté comprendida entre el 17 y el 23 %.

Población : $\begin{cases} p = 0{,}2 \\ q = 0{,}8 \end{cases}$; muestra: n = 50

$$\mu_{\hat{p}} = p = 0{,}2;\ \sigma_{\hat{p}} = \sqrt{\frac{p \cdot q}{n}} = \sqrt{\frac{0{,}2 \cdot 0{,}8}{50}} = 0{,}057$$

$$\hat{p} \sim N(0{,}2; 0{,}057), o\ bien, z = \frac{\hat{p} - 0{,}2}{0{,}057} \sim N(0; 1)$$

a) $p(\hat{p} > 0{,}25) = p\left(z > \frac{0{,}25 - 0{,}2}{0{,}057}\right) = p(z > 0{,}88) = 0{,}1894\ o\ 18{,}94\ \%$

b) $p(\hat{p} < 0{,}22) = p\left(z < \frac{0{,}22 - 0{,}2}{0{,}057}\right) = p(z < 0{,}35) = 0{,}6368\ o\ 63{,}68\ \%$

c) $p(0{,}17 < \hat{p} < 0{,}23) = p\left(\frac{0{,}17 - 0{,}2}{0{,}057} < z < \frac{0{,}23 - 0{,}2}{0{,}057}\right) = p(-0{,}53 < z < 0{,}52) = 2 \cdot 0{,}7091 - 1 = 0{,}4038\ o\ 40{,}38\ \%$

Hay una probabilidad del 40,38 % de que el porcentaje de compradores en estos centros esté comprendido entre el 17 y el 23 %.

13. Encuesta de producto. **Los fabricantes de una nueva bebida estiman que el 60 % de los consumidores la valorarán positivamente. En una muestra aleatoria de 100 consumidores, ¿cuál es la probabilidad de que la valoren de forma positiva, al menos, el 55 %?**

Población : $\begin{cases} p = 0{,}6 \\ q = 0{,}4 \end{cases}$; muestra: n = 100

$$\mu_{\hat{p}} = p = 0{,}6;\ \sigma_{\hat{p}} = \sqrt{\frac{p \cdot q}{n}} = \sqrt{\frac{0{,}6 \cdot 0{,}4}{100}} = 0{,}049$$

$$\hat{p} \sim N(0{,}6; 0{,}049), o\ bien, z = \frac{\hat{p} - 0{,}6}{0{,}049} \sim N(0;1)$$

$$p(\hat{p} \geq 0{,}55) = p\left(z \geq \frac{0{,}55 - 0{,}6}{0{,}049}\right) = p(z \geq -1{,}02) = p(z \leq 1{,}02) = 0{,}8461\ o\ 84{,}61\ \%$$

14. Devoluciones de venta *online*. **Una empresa de venta *online* estima que el 20 % de los pedidos se devuelven por problemas de talla. En una muestra de 50 pedidos, ¿cuál es la probabilidad de que se devuelvan más del 25 % de los pedidos por estos problemas?**

Población : $\begin{cases} p = 0{,}2 \\ q = 0{,}8 \end{cases}$; muestra: n = 50

$$\mu_{\hat{p}} = p = 0{,}2;\ \sigma_{\hat{p}} = \sqrt{\frac{p \cdot q}{n}} = \sqrt{\frac{0{,}2 \cdot 0{,}8}{50}} = 0{,}057$$

$$\hat{p} \sim N(0{,}2; 0{,}057), o\ bien, z = \frac{\hat{p} - 0{,}2}{0{,}057} \sim N(0;1)$$

$$p(\hat{p} > 0{,}25) = p\left(z \geq \frac{0{,}25 - 0{,}2}{0{,}057}\right) = p(z > 0{,}88) = 1 - p(z \leq 0{,}88) = 1 - 0{,}8106\ o\ 81{,}06\ \%$$

15. Aficiones. **Se sabe que el 40 % de los estudiantes de una determinada provincia son aficionados al fútbol. Si se elige una muestra de 200 estudiantes, ¿cuál es la probabilidad de que el porcentaje de aficionados de dicha muestra esté entre el 35 y el 45 %?**

Población : $\begin{cases} p = 0{,}4 \\ q = 0{,}6 \end{cases}$; muestra: n = 200

$$\mu_{\hat{p}} = p = 0{,}4;\ \sigma_{\hat{p}} = \sqrt{\frac{p \cdot q}{n}} = \sqrt{\frac{0{,}4 \cdot 0{,}6}{200}} = 0{,}035$$

$$\hat{p} \sim N(0{,}4;\, 0{,}035),\, o\ bien,\, z = \frac{\hat{p} - 0{,}4}{0{,}035} \sim N(0;1)$$

$$p(0{,}35 < \hat{p} < 0{,}45) = p\left(\frac{0{,}35 - 0{,}40}{0{,}035} < z < \frac{0{,}45 - 0{,}40}{0{,}035}\right) = p(-1{,}43 < z < 1{,}43)$$
$$= 2 \cdot p(z < 01{,}43) - 1 = 0{,}8472\ o\ 84{,}72\ \%$$

Estimación

16. Las ventas (en miles de euros) de ocho establecimientos un día determinado fueron las siguientes:

3 2 5,4 6 3,54 2,43 2 1,75

Encuentra un estimador de las ventas medias.

Un estimador de las ventas medias es la media muestra. Así pues:

$$\bar{x} = \frac{3 + 2 + 5{,}4 + 6 + 3{,}54 + 2{,}43 + 2 + 1{,}75}{8} = 3{,}265$$

Se estima que las ventas medias de este año serán de 3265 €.

17. Las edades de los pacientes en una consulta médica en un día fijado fueron:

16, 15, 13, 18, 19, 13, 20, 16, 15, 22,21 y 23

Halla una estimación puntual de los pacientes mayores de edad.

Los pacientes mayores de edad tienen 18, 19, 20, 22, 21 y 23 años. Una estimación puntual de estos pacientes es la media muestra. Así:

$$\bar{x} = \frac{18 + 19 + 20 + 22 + 21 + 23}{6} = 20{,}5\ años$$

Luego la edad media se estima en 20,5 años.

18. Supóngase que los intervalos [3,7] y [2,8] son dos intervalos de confianza de un parámetro de la población con niveles de confianza del 95 y del 99 %. ¿Qué interpretación se puede dar?

$[3,\, 7]_{95\ \%}\, [2,\, 8]_{99\ \%}$

La probabilidad de que el parámetro poblacional esté comprendido entre 3 y 7 es del 95 %. En cambio, la probabilidad de que esté comprendido entre 2 y 8 es del 99 %.

Por otra parte, se observa cómo el intervalo [2, 8], con un nivel de confianza del 99 %, tiene más amplitud que el intervalo [3, 7] con un nivel de confianza del 95 %.

19. Si [165,182] es un intervalo de confianza para la altura media, en centímetros, de los estudiantes de Bachillerato de un IES a un nivel de confianza del 90 %, interprétalo. Si el nivel de confianza fuera del 99 %, ¿qué le ocurriría al intervalo anterior?

Significa que hay una probabilidad del 90 % de que la altura media de los estudiantes de Bachillerato de ese IES se encuentre entre 165 y 182 centímetros.

Por otro lado, si aumentamos el nivel de confianza al 99 %, entonces la amplitud del intervalo será mayor y, por tanto, se pierde precisión.

Intervalos de confianza para la media μ de una población

20. En una muestra aleatoria simple de 200 trabajadores de una multinacional se encontró que el salario medio anual era de 25 100 €, con una desviación típica de 3000 €. Construye un intervalo de confianza para la media poblacional, es decir, el salario medio de todos los empleados, con un 95 % de nivel de confianza.

Del enunciado deducimos lo siguiente:

Población: ?

Muestra: $\begin{cases} n = 200 \\ \bar{x} = 25\,100\ € \\ \hat{s} = 3000\ € \end{cases}$

Tenemos que construir un intervalo de confianza para la media poblacional μ, con una desviación típica σ desconocida, procedente de una población no normal y de tamaño n = 200.

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

Para un nivel de confianza del 95 %, se obtiene $z_{\alpha/2} = 1{,}96$. Así pues:

$$\left[25\,000 - 1{,}96 \cdot \frac{3000}{\sqrt{200}}, 25\,000 + 1{,}96 \cdot \frac{3000}{\sqrt{200}}\right] = [24\,684{,}22; 25\,515{,}78]_{95\,\%}$$

El salario medio anual de los empleados de esta multinacional oscila entre esos dos importes.

21. Se extrae una muestra de tamaño 36 con una media de 45 y una desviación típica de 12. ¿Es probable que la media poblacional sea 50? Utiliza niveles de confianza del 90, 95 y 99 %.

Población: $\mu = 50$?

Muestra: $\begin{cases} n = 36 \\ \bar{x} = 45 \\ \hat{s} = 12 \end{cases}$

El intervalo de confianza para la media poblacional μ procedente de una población no normal con desviación típica desconocida es:

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

Si el nivel de confianza es del 90 %, entonces $z_{\alpha/2} = 1{,}645$, luego $45 \pm 1{,}645 \cdot \frac{12}{6} = 45 \pm 3{,}29 = [41{,}71; 48{,}39]_{90\%}$.

Si el nivel de confianza es del 95 %, entonces $z_{\alpha/2} = 1{,}96$, luego $45 \pm 1{,}96 \cdot \frac{12}{6} = 45 \pm 3{,}92 = [41{,}08; 48{,}92]_{95\%}$.

Si el nivel de confianza es del 99 %, entonces $z_{\alpha/2} = 2{,}575$, luego $45 \pm 2{,}575 \frac{12}{6} = 45 \pm 5{,}15 = [39{,}85; 50{,}15]_{99\%}$.

Por tanto, $\mu = 50 \in [39{,}85; 50{,}15]$ y, en consecuencia, el nivel de confianza es del 99 %. Sin embargo, no es probable a niveles de confianza del 90 y 95 %.

22. En una muestra aleatoria de tamaño 100 se obtuvo que la media era 25 y la desviación típica 50. Calcula una estimación puntual para µ. Construye un intervalo de confianza para µ con un nivel de confianza del 95 %.

Muestra: $\begin{cases} n = 100 \\ \bar{x} = 25 \\ \hat{s} = 50 \end{cases}$

Una estimación puntual para la media poblacional μ es la media poblacional: $\bar{x} = 25$.

Como para un nivel de significación del 95 % el valor $z_{\alpha/2} = 1{,}96$, entonces el intervalo de confianza para la media poblacional es:

$$\bar{x} \pm z_{\alpha/2} \cdot \frac{\hat{s}}{\sqrt{n}} = 25 \pm 1{,}96 \cdot \frac{50}{10} = 25 \pm 9{,}8 = [15{,}2; 34{,}8]_{95\%}$$

Es decir, a un nivel de significación del 95%, la media poblacional se encontrará entre 15,2 y 34,8.

23. **Valoración de políticos. El barómetro del CIS de un mes determinado ofrecía los siguientes resultados sobre la valoración de dos líderes políticos. Calcula un intervalo de confianza del 95 % para la media de valoraciones recibidas por A y por B.**

Líder	$\overline{x}$	$\hat{s}$	n
A	3,68	2,47	1947
B	2,72	2,35	1351

Un intervalo de confianza del 95 % para la media del político A es:

$$\bar{x}_A \pm z_{\alpha/2} \cdot \frac{\hat{s}_A}{n_A} = 3{,}68 \pm 1{,}96 \cdot \frac{2{,}47}{\sqrt{1947}} = 3{,}68 \pm 0{,}1098 = [3{,}57; 3{,}79]_{95\%}$$

Por su parte, un intervalo de confianza del 95 % para la valoración del político B sería:

$$\bar{x}_B \pm z_{\alpha/2} \cdot \frac{\hat{s}_B}{n_B} = 2{,}72 \pm 1{,}96 \cdot \frac{2{,}35}{\sqrt{1351}} = 2{,}72 \pm 0{,}1253 = [2{,}59; 2{,}85]_{95\%}$$

24. **Producción de naranjas. La producción, en kilogramos, de naranjas por naranjo en una población mallorquina sigue una distribución normal con una desviación típica de 2 kg. Se toma una muestra de 10 naranjos cuya producción en kilogramos es:**

30, 25, 4, 70, 45, 60, 21, 32, 9 y 47

Calcula un intervalo de confianza al 97 % para estimar la producción media de naranjas por árbol.

$$\begin{matrix} Población: \\ N(\mu, \sigma) \end{matrix} \{ \sigma = 2\ kg \qquad Muestra: \begin{cases} n = 10 \\ \bar{x}\ ? \end{cases}$$

$$\bar{x} = \frac{30 + 25 + 4 + 70 + 45 + 60 + 21 + 32 + 9 + 47}{10} = 34{,}3\ kg$$

Si $1 - \alpha = 0{,}97 \rightarrow \alpha = 0{,}03 \rightarrow \frac{\alpha}{2} = 0{,}015 \rightarrow 1 - \frac{\alpha}{2} = 0{,}985 \rightarrow z_{\frac{\alpha}{2}} = 2{,}17$

Un intervalo de confianza para la producción media de naranjas por árbol es:

$$\bar{x} \pm z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} = 34{,}3 \pm 2{,}17 \cdot \frac{2}{\sqrt{10}} = 34{,}3 \pm 1{,}372 = [32{,}928; 35{,}672]_{97\%}$$

Al 97 % de confianza, podemos afirmar que la producción media por naranjo está comprendida entre 32,928 y 35,672 kg.

Intervalo de confianza para la proporción p de una población

25. Personas zurdas. En una muestra de 100 personas se encontró que 8 eran zurdas. Estima la verdadera proporción de la población hallando una estimación puntual y por intervalo de confianza al 90 %. Después calcula los intervalos de confianza al 95 y 99 %, y compáralos con el inicial.

Datos de la encuesta: $\begin{cases} n = 100 \\ \hat{p} = \frac{8}{100} = 0{,}08 \\ \hat{q} = 0{,}92 \end{cases}$

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

Una estimación puntual para la proporción de personas zurdas, p, sería:

$p = \hat{p} = \frac{8}{100} = 0{,}08$ u 8 %

Por consiguiente, se estima que el 8% de la población son personas zurdas.

Un intervalo de confianza del 90 % para la proporción de personas zurdas, p, es:

$$\hat{p} \pm z_{\alpha/2} \cdot \sqrt{\frac{\hat{p} \cdot \hat{q}}{n}} = 0{,}08 \pm 1{,}645 \cdot \sqrt{\frac{0{,}08 \cdot 0{,}92}{100}} = 0{,}08 \pm 0{,}0446 = [0{,}0354; 0{,}1246]_{90\%}$$

A un nivel de confianza del 90 %, la proporción de personas zurdas está comprendía entre el 3,54 % y el 12,46 % del total.

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

$$\hat{p} \pm z_{\alpha/2} \cdot \sqrt{\frac{\hat{p} \cdot \hat{q}}{n}} = 0{,}08 \pm 1{,}96 \cdot \sqrt{\frac{0{,}08 \cdot 0{,}92}{100}} = 0{,}08 \pm 0{,}0531 = [0{,}0269; 0{,}1331]_{95\%}$$

A un nivel de confianza del 95 %, la proporción de personas zurdas está comprendía entre el 2,69 % y el 13,31 % del total.

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

$$\hat{p} \pm z_{\alpha/2} \cdot \sqrt{\frac{\hat{p} \cdot \hat{q}}{n}} = 0{,}08 \pm 2{,}575 \cdot \sqrt{\frac{0{,}08 \cdot 0{,}92}{100}} = 0{,}08 \pm 0{,}0699 = [0{,}0101; 0{,}1499]_{99\%}$$

A un nivel de confianza del 99 %, la proporción de personas zurdas está comprendía entre el 1,01 % y el 14,99 % del total.

Podemos ver que, con el aumento del nivel de confianza, la amplitud del intervalo obtenido se incrementa y pierde precisión.

26. **En una determinada ciudad se desea conocer la proporción actual de hogares con más de un automóvil. Para ello, se realiza una muestra de 300 hogares y resulta que en 120 hay más de un automóvil. Construye un intervalo de confianza para la proporción de hogares con esa condición, a un nivel de confianza del 95 %.**

n = 300

$\hat{p} = \frac{120}{300} = 0{,}4$

$\hat{q} = 0{,}6$

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

Un intervalo de confianza para la proporción poblacional p a un nivel de confianza del 95 % es:

$$\hat{p} \pm z_{\alpha/2} \cdot \sqrt{\frac{\hat{p} \cdot \hat{q}}{n}} = 0{,}4 \pm 1{,}96 \cdot \sqrt{\frac{0{,}4 \cdot 0{,}6}{300}} = 0{,}4 \pm 0{,}0554 = [0{,}3446; 0{,}4554]_{95\%}$$

La proporción de hogares en esta ciudad con más de un automóvil oscila entre el 34,46 y el 45,54 %, para un nivel de confianza del 95 %.

Relación entre tamaño muestral, nivel de confianza y error de la estimación

27. Encuesta de satisfacción. **Calcula el tamaño muestral necesario para realizar una encuesta a los universitarios granadinos sobre su satisfacción en los estudios que cursan, con un nivel de confianza del 97,5 % y un error máximo muestral del 2 %. Considera p = q.**

Si el nivel de confianza es del 97,5 %, entonces:

Si $1 - \alpha = 0{,}975 \rightarrow \alpha = 0{,}025 \rightarrow \frac{\alpha}{2} = 0{,}0125 \rightarrow 1 - \frac{\alpha}{2} = 0{,}9875 \rightarrow z_{\frac{\alpha}{2}} = 2{,}24$

El tamaño muestral es la estimación de la proporción poblacional viene dado por:

$$n = \frac{z^2_{\alpha/2} \cdot \hat{p} \cdot \hat{q}}{E^2}$$

Como $\hat{p} = \hat{q} = 0{,}5$, entonces:

$$n = 3136$$

El tamaño muestral para esta encuesta de satisfacción a los estudiantes granadinos debe ser de, al menos, 3136 encuestados con el fin de que el error máximo muestral sea del 2 % para un nivel de confianza del 97,5 %.

28. Vacuna. **El 25 % de una población ha sido vacunada contra una enfermedad. Determina el tamaño mínimo que debería tener una muestra de dicha población para que, con un nivel de confianza del 95 %, la proporción muestral y la proporción poblacional no difieran en más del 2 %.**

Muestra: $\begin{cases} n\,? \\ \hat{p} = 0{,}25 \\ \hat{q} = 0{,}75 \end{cases}$

Nivel de confianza: $1 - \alpha = 0{,}95 \rightarrow z_{\frac{\alpha}{2}} = 1{,}96$

El tamaño muestral viene dado por:

$$n = \frac{z_{\alpha/2}^2 \cdot \hat{p} \cdot \hat{q}}{E^2} = \frac{(1{,}96)^2 \cdot 0{,}25 \cdot 0{,}75}{(0{,}02)^2} = 1800{,}75$$

El tamaño mínimo requerido es de 1801 personas con el fin de que, el error muestral sea como máximo del 2 %, y a un nivel de confianza del 95 %.

29. Viajes. **Un grupo de agencias de viajes quiere determinar la proporción de españoles que ha viajado alguna vez a Europa y desea estar seguro de que el error de muestreo no sea mayor del 3 %. Suponiendo que el investigador no sepa el valor real de la proporción de la población, ¿qué tamaño de la muestra es necesario realizar para tener un nivel de confianza del 95 %?**

Nivel de confianza: $1 - \alpha = 0{,}95 \rightarrow z_{\frac{\alpha}{2}} = 1{,}96$

El tamaño muestral, n, en la estimación de la proporción poblacional, p, viene dado por:

$$n = \frac{z_{\alpha/2}^2 \cdot \hat{p} \cdot \hat{q}}{E^2}$$

Como no disponemos de información acerca de la proporción muestral, $\hat{p}$, consideramos el caso más desfavorable o de mayor varianza, que se da cuando $\hat{p} = \hat{q} = 0{,}5$. Así pues:

$$n = \frac{z_{\alpha/2}^2 \cdot \hat{p} \cdot \hat{q}}{E^2} = \frac{(1{,}96)^2 \cdot 0{,}5 \cdot 0{,}5}{(0{,}03)^2} = 1067{,}11$$

Por tanto, la muestra tiene que ser de, al menos, 1068 personas para que el error del muestreo no supere el 3 %.

30. Una muestra aleatoria extraída de una población normal de varianza igual a 100 tiene de media 160. Si se quiere tener una confianza del 95 % de que el error máximo de la estimación sea del 1,2 %, ¿de qué tamaño tiene que ser la muestra?

$$\begin{matrix} Población: \\ N(\mu,\sigma) \end{matrix} \begin{cases} \sigma^2 = 100 \\ \mu = 160 \end{cases}$$

Nivel de confianza: $1 - \alpha = 0{,}95 \rightarrow z_{\frac{\alpha}{2}} = 1{,}96$. Error máximo de la solución: E = 0,012.

El tamaño muestral necesario para la estimación de una población normal con σ conocida es:

$$n = \left(\frac{z_{\alpha/2} \cdot \sigma}{E}\right) = \frac{1{,}96 \cdot 10}{0{,}012} = 2667777{,}778$$

La muestra debería ser de tamaño mínimo: 2 667 778. No parece, por tanto, muy razonable el planteamiento de este problema.

Contrastes de hipótesis

31. Black Friday. Según una asociación de empresarios de ventas por internet, el número medio de devoluciones en Black Friday es 200. Por experiencias anteriores se acepta que el número de devoluciones en esa fecha sigue una distribución normal con una desviación típica de 25. Para contrastar esta afirmación y decidir si es superior, se toma una muestra de 800 ventas y se produjo una media de 210 devoluciones. Contrasta la afirmación realizada por esta asociación a un nivel de significación del 10 %.

Se trata del siguiente contraste unilateral para la media poblacional μ:

$$\begin{cases} H_0: \mu = 200 \\ H_1: \mu > 200 \end{cases}$$

Población: $\{\sigma = 25$

Muestra: $\begin{cases} n = 800 \\ \bar{x} = 210 \end{cases}$

Como el nivel de significación es del 10 %, entonces $z_{\frac{\alpha}{2}} = 1{,}645$.

Por otro lado, el estadístico de contraste es:

$$z_0 = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} = \frac{210 - 200}{25/\sqrt{800}} = 11{,}31$$

Como $z_0 = 11{,}31 > z_{\frac{\alpha}{2}} = 1{,}645$, se rechaza la hipótesis nula, es decir, que el número medio de devoluciones durante el Black Friday sea de menos de 200, como afirma la asociación de empresarios.

32. Sandías. Se supone que el peso medio de las sandías de una variedad determinada sigue una distribución normal con una desviación típica de 1 kg. Se toma una muestra aleatoria de 100 sandías y se observa que el peso medio es de 6 kg.

a) Calcula un intervalo de confianza al 95 % para el peso medio de esa variedad de sandía.

b) ¿Puede aceptarse la hipótesis de que el verdadero peso medio de las sandías es de 5 kg, frente a que sea diferente, con un nivel de significación del 5 %?

$\underset{N(\mu,\sigma)}{Población}: \begin{cases} \sigma = 1\ kg \\ \mu = ? \end{cases}$

Muestra: $\begin{cases} n = 100\ sandías \\ \bar{x} = 6\ kg \end{cases}$

a) Nivel de confianza: si $1 - \alpha = 0{,}95 \rightarrow z_{\frac{\alpha}{2}} = 1{,}96$.

El intervalo de confianza para el peso medio de esta variedad de sandía es:

$$\bar{x} \pm z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} = 6 \pm 1{,}96 \cdot \frac{1}{10} = 6 \pm 0{,}196 = [5{,}804; 6{,}196]_{95\ \%}$$

Con una confianza del 95 % se puede estimar que el peso medio de las sandías se encontrará entre 5,804 y 6,196 kg.

b) Se trata de un contraste bilateral para el peso medio de estas sandías. Es decir:

$\begin{cases} H_0: \mu = 5\ kg \\ H_1: \mu \neq 5\ kg \end{cases}$

El estadístico de contraste es:

$$z_0 = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} = \frac{6 - 5}{1/10} = 10$$

Como $z_0 = 10 > z_{\frac{\alpha}{2}} = 1{,}96$, se rechaza la hipótesis de que el peso medio de las sandías es de 5 kg para un nivel de significación del 5 %. Podríamos haberlo comprobado de la siguiente manera:

Como $\mu_0 = 5 \notin [5{,}804; 6{,}196]_{95\ \%}$, se rechaza la hipótesis.

Actividades finales

Tipos de muestreo. Teorema central del límite

33. Según información facilitada por el Instituto Nacional de Estadística (INE), la población mayor de 18 años en Aragón se distribuye de la siguiente manera conforme al último censo:

Provincia	Nº de habitantes
Huesca	223 995
Teruel	134 259
Zaragoza	973 684
Total	1 331 938

Distribuye una muestra de 2000 entrevistas mediante afijación proporcional al tamaño demográfico de la provincia de Aragón.

Tamaño de la población: N = 1 331 938

Tamaño de la muestra: n = 2000

Para que la afijación sea proporcional al tamaño de cada provincia, si n_1, n_2 y n_3 representan, respectivamente, los tamaños muestrales de Huesca, Teruel y Zaragoza. Entonces:

$$\frac{2000}{1331938} = \frac{n_1}{223995} = \frac{n_2}{134259} = \frac{n_3}{973684}$$

Por tanto: $n_1 = 336$, $n_2 = 202$, $n_3 = 1462$

Habría que realizar 336 entrevistas en Huesca, 202 en Teruel y 1462 en Zaragoza.

34. Se desea diseñar un muestreo para conocer la opinión de los 150 profesores de un centro de Secundaria. De los 6 departamentos existentes, se pretende seleccionar 2 y, de estos, elegir una muestra de 20 profesores. ¿Qué tipo de muestreo es el más adecuado? Razónalo.

Se dividen en los 150 profesores en los 6 estratos que determinan los departamentos. Después, por ejemplo, mediante un muestreo sistemático seleccionamos 2 departamentos. Y, finalmente, mediante afijación proporcional al número de profesores de esos dos departamentos, seleccionamos la muestra de 20 profesores.

35. **Una grúa funciona con cuatro motores eléctricos de forma conjunta. Para que la grúa funcione correctamente, los cuatro motores deben generar, al menos, 380 CV. La empresa que fabrica los motores afirma que sigue la ley normal de media 100 CV y una desviación típica de 10 CV. A la encargada de mantenimiento de esta gama de motores le preocupa que los motores no sean lo bastante potentes para que la grúa funcione correctamente. ¿Es razonable la preocupación de la encargada de mantenimiento?**

Sea x_i = «potencia, en CV, del motor i», con i = 1, 2, 3, 4

$$x_i \sim N(100, 10)$$

Entonces, la suma de los cuatro motores, según el teorema central del límite, verifica:

$$S = x_1 + x_2 + x_3 + x_4 = \sum_{i=1}^{4} x_i$$

$$\left.\begin{matrix} \mu_S = 4 \cdot 100 = 400 \\ \sigma_S = 10 \cdot \sqrt{4} = 20 \end{matrix}\right\} \to S \sim N(400, 20)$$

O bien, $z = \frac{S-400}{20} \sim N(0,1)$

$$p(S > 380) = p\left(z > \frac{380 - 400}{20}\right) = p(z > -1) = p(z < 1) = 0{,}8413 \; o \; 84{,}13\ \%$$

Distribución de la media y proporción muestrales

36. **La edad media de los alumnos de un club de tenis es de 16,3 años y una desviación típica de 3,5 años. Si se selecciona una muestra de 40 alumnos, ¿cuál es la probabilidad de que la edad media de esta muestra esté comprendida entre los 16 y 17 años?**

Población: $\begin{cases} \mu = 16{,}3 \text{ años} \\ \sigma = 3{,}5 \text{ años} \end{cases}$

Muestra: n = 40 alumnos

La distribución muestral de la media es:

$$\mu_{\bar{x}} = \mu = 16{,}3; \; \sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = 0{,}55$$

$\bar{x} \sim N(16{,}3; 0{,}55)$, o bien: $z = \frac{\bar{x} - 16{,}3}{0{,}55} \sim N(0,1)$

$$p(16 < \bar{x} < 17) = p\left(\frac{16 - 16{,}3}{0{,}55} < z < \frac{17 - 16{,}3}{0{,}55}\right) = p(-0{,}55 < z < 1{,}27)$$
$$= 0{,}6058 \; o \; 60{,}58\ \%$$

37. **La edad de los miembros de un colegio profesional sigue una distribución N(μ,σ). Se conoce que la distribución de las medias de las edades en una muestra de 36 colegiados tiene de media 52 años y una desviación de 0,5 años. Determina la media y la desviación típica de la edad de los miembros del colegio.**

Población: $N(\mu, \sigma)$

Muestra: $\begin{cases} \bar{x} = 52 \text{ años} \\ \hat{s} = 0{,}5 \text{ años} \\ n = 36 \end{cases}$

$\mu_{\bar{x}} = \mu = \bar{x} = 52;\ \sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \hat{s} = 0{,}5$, luego $\sigma = 3$.

Por consiguiente, la edad media de los miembros es de 52 años, mientras que la desviación típica se observa que es de 3 años.

38. **El peso de los estudiantes de Primaria de un centro escolar se distribuye según una ley normal de media 38 kg y una desviación típica de 3 kg. Se selecciona al azar una muestra de 36 estudiantes y se calcula su peso medio. ¿Cuál es la probabilidad de que se encuentre entre los 37 y 39 kg?**

Población: $N(\mu, \sigma)$

Muestra: $\begin{cases} \mu = 38\ kg \\ \sigma = 3\ kg \\ n = 36 \end{cases}$

Así pues, la distribución de la media muestra verifica:

$$\mu_{\bar{x}} = \mu = 38;\ \sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{3}{6} = 0{,}5$$

$\bar{x} \sim N(38; 0{,}5)$, o bien: $z = \frac{\bar{x}-38}{0{,}5} \sim N(0{,}1)$

$$p(37 < \bar{x} < 39) = p\left(\frac{37-38}{0{,}5} < z < \frac{39-38}{0{,}5}\right) = p(-2 < z < 2) = 0{,}9544\ o\ 95{,}44\ \%$$

39. **En el proceso de producción de determinadas piezas se sabe que la desviación típica de su peso es de 25 g. ¿Cuál debe ser el peso medio de las piezas, si en una muestra aleatoria de 50 piezas la probabilidad de que el peso medio sea superior a 250 g es de 0,95?**

Población: $N(\mu, \sigma)$

Muestra: $\begin{cases} \mu = ? \\ \sigma = 25\ g \\ n = 50 \end{cases}$

Así pues, la distribución de la media muestra verifica:

$$\mu_{\bar{x}} = \mu;\ \sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{25}{\sqrt{50}} = 3{,}54$$

$\bar{x} \sim N(\mu; 3{,}54)$, o bien: $z = \frac{\bar{x}-\mu}{3{,}54} \sim N(0{,}1)$

Como en la muestra de 50 piezas la probabilidad de que el peso medio $\bar{x}$ sea superior a 250 g es de 0,95, entonces:

$$p(\bar{x} > 250) = 0{,}95 \rightarrow p(\bar{x} \leq 250) = 0{,}05 \rightarrow p\left(\frac{\bar{x}-\mu}{3{,}54} \leq \frac{250-\mu}{3{,}54}\right) = 0{,}05 \rightarrow p\left(z \leq \frac{250-\mu}{3{,}54}\right) = 0{,}05$$
$$\rightarrow \mu = 252{,}82 \approx 253$$

El peso medio de las piezas debe ser de aproximadamente 253 gramos.

40. En una comunidad autónoma de 9 millones de habitantes, la edad media es de 41 años y la desviación típica es de 12 años. ¿Qué proporción de las muestras de 1000 personas que se pueden elegir de esta comunidad tiene una edad media inferior a 35 años? ¿Y superior a 40 años?

Población: $N(\mu, \sigma)$

Muestra: $\begin{cases} \mu = 41\ años \\ \sigma = 12\ años \\ n = 1000\ personas \end{cases}$

Así pues, la distribución de la media muestra verifica:

$$\mu_{\bar{x}} = \mu = 41;\ \sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{12}{\sqrt{1000}} = 0{,}38$$

$\bar{x} \sim N(41; 0{,}38)$, o bien: $z = \frac{\bar{x}-41}{0{,}38} \sim N(0{,}1)$

$$p(\bar{x} < 35) = p\left(z \leq \frac{35-41}{0{,}38}\right) = p(z < -15{,}79) = 0$$

Por consiguiente, ninguna muestra tendrá una edad media inferior a 35 años.

$$p(\bar{x} > 40) = p\left(z > \frac{40-41}{0{,}38}\right) = p(z > -2{,}63) = (z < 2{,}63) = 0{,}9957\ o\ 99{,}57\ \%$$

El 99,57 % de las muestras tendrán una edad media superior a 40 años.

41. Se extrae una muestra aleatoria simple de tamaño 300 de una población en la que p = 0,6. Halla la probabilidad de que la proporción muestral:

a) sea inferior a 0,55;

b) esté entre 0,55 y 0,65.

Población: $\begin{cases} p = 0{,}6 \\ q = 0{,}4 \\ n = 300 \end{cases}$

La distribución muestral de la proporción $\hat{p}$ viene dada por:

$$\mu_{\hat{p}} = p = 0{,}6, \sigma_{\hat{p}} = \sqrt{\frac{p \cdot q}{n}} = \sqrt{\frac{0{,}6 \cdot 0{,}4}{300}} = 0{,}028$$

Luego: $\hat{p} \sim N(0{,}6; 0{,}028)$, o bien: $z = \frac{\hat{p}-0{,}6}{0{,}028} \sim N(0{,}1)$

a) $p(\hat{p} < 0{,}55) = p\left(z < \frac{0{,}55-0{,}6}{0{,}028}\right) = p(z < -1{,}79) = 0{,}0367 \ o \ 3{,}67\ \%$

b) $p(0{,}55 < \hat{p} < 0{,}65) = p\left(\frac{0{,}55-0{,}6}{0{,}028} < z < \frac{0{,}65-0{,}6}{0{,}028}\right) = p(-1{,}79 < z < 1{,}79) = 0{,}9266 \ o \ 92{,}66\ \%$

42. De los 1200 automóviles vendidos al año por un concesionario, solo 30 tuvieron que acudir al taller por diversos problemas durante el primer año. En una muestra de 60 automóviles, ¿cuál es la probabilidad de que, como mucho, dos de ellos hayan ido al taller durante el primer año por algún problema?

Población: $\begin{cases} p = 0{,}025 \\ q = 0{,}975 \\ n = 60 \end{cases}$

La distribución muestral de la proporción $\hat{p}$ viene dada por:

$$\mu_{\hat{p}} = p = 0{,}025, \sigma_{\hat{p}} = \sqrt{\frac{p \cdot q}{n}} = \sqrt{\frac{0{,}025 \cdot 0{,}975}{60}} = 0{,}02$$

Luego: $\hat{p} \sim N(0{,}025; 0{,}02)$, o bien: $z = \frac{\hat{p}-0{,}025}{0{,}02} \sim N(0{,}1)$

a) $p\left(\hat{p} \le \frac{2}{60}\right) = p\left(z \le \frac{2/60-0{,}025}{0{,}02}\right) = p(z \le 0{,}42) = 0{,}6628 \ o \ 66{,}28\ \%$

La probabilidad de que, en una muestra de 60 automóviles como mucho, 2 de ellos hayan ido al taller durante el primer año por algún problema es del 66,28 %.

Estimación por intervalos de confianza

43. **En una muestra aleatoria de tamaño 36 se sabe que la media es 30 y la desviación típica es 8. Halla un intervalo de confianza al 95 % para la media poblacional. Si la muestra es de tamaño 100, ¿cómo se modifica el intervalo anterior? Interprétalo.**

Población: $\begin{cases} \mu = ? \\ \sigma = ? \end{cases}$

Muestra: $\begin{cases} n = 36 \\ \bar{x} = 30 \\ \hat{s} = 8 \end{cases}$

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

El intervalo de confianza para la media poblacional μ con σ desconocida es:

$$\bar{x} \pm z_{\frac{\alpha}{2}} \cdot \frac{\hat{s}}{\sqrt{n}} = 30 \pm 1{,}96 \cdot \frac{8}{6} = 30 \pm 2{,}61 = [27{,}39; 32{,}61]_{95\ \%}$$

Por otra parte, si n = 100, entonces:

$$\bar{x} \pm z_{\frac{\alpha}{2}} \cdot \frac{\hat{s}}{\sqrt{n}} = 30 \pm 1{,}96 \cdot \frac{8}{10} = 30 \pm 1{,}57 = [28{,}43; 31{,}57]_{95\ \%}$$

Al aumentar el tamaño muestral, se reduce la amplitud del intervalo de confianza y, por tanto, se gana en precisión con respecto al a estimación de μ.

44. **El tiempo de espera para pagar en una caja de un supermercado se distribuye según una ley normal con una desviación típica de 2,5 minutos. Una muestra aleatoria de 25 clientes reflejó que, por término medio, esperaban 10,6 minutos.**

a) Construye un intervalo de confianza al 98 % para el tiempo medio de espera en una caja de dicho supermercado.

b) ¿Es razonable compartir la información del supermercado de que sus clientes, por término medio, no esperan en la caja más de 9 minutos?

Población, N(μ, σ): $\begin{cases} \mu = ? \\ \sigma = 2{,}5\ min \end{cases}$

Muestra: $\begin{cases} n = 25 \\ \bar{x} = 10{,}6\ min \end{cases}$

$$1 - \alpha = 0{,}98 \rightarrow z_{\frac{\alpha}{2}} = 2{,}33$$

a) El intervalo para la media muestral poblacional μ es:

$$\bar{x} \pm z_{\frac{\alpha}{2}} \cdot \frac{\hat{s}}{\sqrt{n}} = 10{,}6 \pm 2{,}33 \cdot \frac{2{,}5}{5} = 10{,}6 \pm 1{,}165 = [9{,}4; 11{,}8]_{98\,\%}$$

El tiempo medio de espera en una caja de este supermercado se encuentra entre 9,4 y 11,8 minutos, con un nivel de confianza del 98 %.

b) Como $\mu < 9$ no se encuentra en el Intervalo obtenido en el apartado anterior, parece razonable rechazar la información del supermercado al 98% de confianza.

45. **Los ingresos brutos anuales de los hogares de un determinado municipio siguen una distribución normal con una desviación típica de 4000 €. Determina el número de hogares que se debe elegir en la muestra, para que no exceda en 300 € el error de la estimación de los ingresos medios, y de modo que la probabilidad de superar este límite sea de 0,05.**

Población, N(μ, σ): $\begin{cases} \mu = ? \\ \sigma = 4000\text{ €} \end{cases}$

Muestra: $\begin{cases} n = ? \\ \bar{x} = ? \end{cases}$

El intervalo de confianza para la media poblacional μ es:

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

Según el enunciado, tenemos:

$$|\bar{x} - \mu| = 300 \rightarrow z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}} = 300$$

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

Luego: $1{,}96 \cdot \frac{4000}{\sqrt{n}} = 300 \rightarrow n = 682{,}95$

En la muestra se deben seleccionar 683 hogares para que se verifiquen estos estudios.

46. **En una encuesta realizada a 1970 personas sobre la valoración (en una escala de 1 a 10) de un determinado partido político, se obtuvo una valoración media de 2,18 y una desviación típica de 1,39. Halla un intervalo de confianza al 95 % para la valoración media de la población española respecto a ese partido político.**

Población = ?

Muestra: $\begin{cases} n = 1970 \\ \bar{x} = 2{,}18 \\ \hat{s} = 1{,}39 \end{cases}$

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

El intervalo de confianza para la media poblacional μ es:

$$\mu \epsilon \left[\bar{x} - z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}, \bar{x} + z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}\right] = \left[2{,}18 - 1{,}96 \cdot \frac{1{,}39}{\sqrt{1970}}, 2{,}18 + 1{,}96 \cdot \frac{1{,}39}{\sqrt{1970}}\right] = [2{,}12; 2{,}24]_{95\,\%}$$

La valoración media ideológica de la población española a ese partido político lo sitúa entre el 2,12 y el 2,24 con una confianza del 95 %.

47. La altura, en metros, de los jugadores de una liga de baloncesto sigue una distribución normal de media desconocida y una desviación típica de 0,3 m. Se sabe que un intervalo de confianza del 95 % para estimar la altura media es [1,9112; 2,0288]. Halla el tamaño de la muestra seleccionada y la altura media de los jugadores que la componían.

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

Población, N(μ, σ): $\begin{cases} \mu = ? \\ \sigma = \ 0{,}3\ m \end{cases}$

Muestra: $\begin{cases} n = ? \\ \bar{x} = ? \end{cases}$

El intervalo de confianza para la altura media μ es:

$$\mu \epsilon \left[\bar{x} - z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}, \bar{x} + z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}\right] = [1{,}9112; 2{,}0288]_{95\,\%}$$

La longitud del intervalo es: $L = 2{,}0288 - 1{,}9112 = 0{,}1176$

$$L = 2 \cdot z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}} = 0{,}1176 \rightarrow n = 100$$

La muestra se compone de 100 jugadores.

Por otro lado, por ejemplo:

$$\bar{x} - z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}} = \bar{x} - 1{,}96 \cdot \frac{0{,}3}{10} = 1{,}9112 \rightarrow \bar{x} = 1{,}97$$

La altura media de la muestra seleccionada es de 1,97 metros.

48. **Cada vez son más las empresas que acostumbran, cuando se intenta contactar por teléfono con ellas, a responder de forma automática seguida de música. En una muestra realizada a 400 personas sobre qué opinión tienen de este tipo de actuación, el 72 % responde que les parece irritable tal procedimiento y están muy descontentas con las empresas. Construye un intervalo de confianza al 95 % para la proporción de personas de la población que les irrita tal procedimiento.**

Población: $\begin{cases} p = 0{,}72 \\ q = 0{,}28 \\ n = 400 \end{cases}$

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

El intervalo de confianza para la proporción $\hat{p}$ viene dado por:

$$\hat{p} \pm z_{\frac{\alpha}{2}} \cdot \sqrt{\frac{p \cdot q}{n}} = 0{,}72 \pm 1{,}96 \cdot \sqrt{\frac{0{,}72 \cdot 0{,}28}{400}} = [0{,}676; 0{,}764]_{95\ \%}$$

Se estima que, entre el 67,6 y el 76,4 % de la población están muy descontentos con estas empresas y les irrita tal procedimiento, para un nivel de confianza del 95 %.

49. **Se sabe que los estudiantes de una comunidad autónoma duermen un número de horas diarias que se distribuye según una ley normal de media μ horas y una desviación típica de 1,8 horas. A partir de una muestra de 64 estudiantes se ha obtenido el siguiente intervalo de confianza [7,26; 8,14] para la media de la población. Determina el nivel de confianza con el que se ha construido dicho intervalo.**

Población, $N(\mu, \sigma)$: $\begin{cases} \mu = ? \\ \sigma = 1{,}8\ h \end{cases}$

Muestra: $n = 64\ estudiantes$

El intervalo de confianza para la altura media μ con σ conocida es: [7,26; 8,14]. Como el intervalo es:

$\left[\bar{x} - z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}, \bar{x} + z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}\right] = \left[\bar{x} - z_{\frac{\alpha}{2}} \cdot \frac{1{,}8}{8}, \bar{x} + z_{\frac{\alpha}{2}} \cdot \frac{1{,}8}{8}\right]$, entonces:

La longitud del intervalo es:

$$L = 2 \cdot z_{\frac{\alpha}{2}} \cdot \frac{1{,}8}{8} \rightarrow 0{,}45 z_{\frac{\alpha}{2}} = 0{,}88 \rightarrow z_{\frac{\alpha}{2}} = 1{,}96$$

Por consiguiente, el nivel de confianza es del 95 %.

Tamaño muestral, nivel de confianza y error de la estimación

50. **Para conocer la aceptación de una nueva bebida, se ha tomado una muestra de 200 personas, de las cuales solo les ha gustado a 50.**

a) Construye un intervalo de confianza para la proporción poblacional de aceptación de dicha bebida con una confianza del 90 %.

b) ¿De cuántas personas debería constar la muestra para estimar dicha proporción poblacional con un error máximo de estimación del 5 %?

Población: $\begin{cases} p = 0{,}25 \\ q = 0{,}75 \\ n = 200 \end{cases}$

Si $1 - \alpha = 0{,}90 \to z_{\frac{\alpha}{2}} = 1{,}645$

a) El intervalo de confianza para la proporción $\hat{p}$ viene dado por:

$$\hat{p} \pm z_{\frac{\alpha}{2}} \cdot \sqrt{\frac{p \cdot q}{n}} = 0{,}25 \pm 1{,}645 \cdot \sqrt{\frac{0{,}25 \cdot 0{,}75}{200}} = [0{,}1996; 0{,}3004]_{90\,\%}$$

b) El tamaño de la muestra n es la estimación de una proporción, al mismo nivel de confianza y suponiendo el caso más desfavorable (es decir, asumiendo que: p = q = 0,5). Así pues:

$$n = \frac{z_{\alpha/2}^2 \cdot \hat{p} \cdot \hat{q}}{E^2} = \frac{1{,}645^2 \cdot 0{,}5 \cdot 0{,}5}{0{,}05^2} = 270{,}6$$

La muestra debería constar de, al menos, 271 personas para que el error máximo de la estimación sea del 5 %.

51. **La edad media de los participantes en unas pruebas de acceso a la Universidad en una provincia es de 18,1 años y una desviación típica de 0,6 años.**

a) Si se elige al azar una muestra de 100 estudiantes, ¿cuál es la probabilidad de que la edad media de la muestra esté comprendida entre 17,9 y 18,3 años?

b) ¿De qué tamaño habría que elegir la muestra para que su media esté comprendida entre 17,9 y 18,3 con una probabilidad del 99,5 %?

a) Población: N(μ, σ): $\begin{cases} \mu = 18{,}1 \text{ años} \\ \sigma = 0{,}6 \text{ años} \\ n = 100 \text{ estudiantes} \end{cases}$

La distribución de la media muestral es:

$$\mu_{\bar{x}} = \mu = 18{,}1; \; \sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{0{,}6}{10} = 0{,}06$$

$\bar{x} \sim N(18{,}1; 0{,}06)$, o bien: $z = \frac{\bar{x}-18{,}1}{0{,}06} \sim N(0{,}1)$

$$p(17{,}9 < \bar{x} < 18{,}3) = p\left(\frac{17{,}9 - 18{,}1}{0{,}06} < z < \frac{18{,}3 - 18{,}1}{0{,}06}\right) = p(-3{,}33 < z < 3{,}33) = 1$$

Prácticamente hay un 100 % de probabilidades de que la edad media se encuentre dentro de ese intervalo.

b) En este caso se desconoce n, luego:

$$p(17{,}9 < \bar{x} < 18{,}3) = 0{,}995$$

Como $\bar{x} \sim N(18{,}1; \frac{0{,}6}{\sqrt{n}})$, entonces $z = \frac{\bar{x}-18{,}1}{\frac{0{,}6}{\sqrt{n}}}$

Por tanto:

$$p\left(-0{,}33\sqrt{n} < z < -0{,}33\sqrt{n}\right) = 0{,}995 \to n \approx 73$$

52. En un estudio sobre las cualidades nutricionales de las llamadas comidas rápidas se ha medido el contenido en grasa de 35 hamburguesas elegidas al azar en un restaurante, observándose un contenido medio de 30,2 g. Se supone que el contenido en grasa de una hamburguesa es una variable aleatoria con una distribución normal de una desviación típica de 3,8 g.

a) Halla un intervalo de confianza del 95 % para el contenido medio de grasa de las hamburguesas servidas en ese restaurante.

b) ¿Cuántas hamburguesas deberían haberse elegido en la muestra para que, con un nivel de confianza del 90 %, se pudiera estimar el contenido medio de grasa con un error inferior a 0,5 g?

a) Población, N(μ, σ): $\begin{cases} \mu = ? \\ \sigma = 3{,}5\ g \end{cases}$

Muestra: $\begin{cases} n = 35 \\ \bar{x} = 30{,}2\ g \end{cases}$

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

El intervalo de confianza para la altura media μ es:

$$\mu \epsilon \left[\bar{x} - z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}, \bar{x} + z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}\right] = [29{,}04; 31{,}36]_{95\,\%}$$

b) Si $1 - \alpha = 0{,}90 \to z_{\frac{\alpha}{2}} = 1{,}645$

El tamaño de la muestra viene dado en este caso por:

$$n = \frac{z_{\alpha/2}^2 \cdot \sigma^2}{E^2} = \frac{1{,}645^2 \cdot 3{,}5 \cdot 3{,}5}{0{,}5^2} = 132{,}60$$

Se deberían elegir 133 hamburguesas con el fin de que el error muestral sea inferior a 0,5 g.

53. Con el fin de estimar la proporción de menores varones condenados en el año 2022, se ha realizado un estudio con una muestra de 500 menores condenados (de 14 a 17 años), de los cuales 405 eran varones.

a) Construye un intervalo de confianza para estimar la población de menores varones condenados con un nivel de confianza del 96,5 %.

b) ¿Cuál es el error máximo cometido en la estimación?

c) Según datos del INE, en el año 2022 el 81 % de los menores condenados eran varones. ¿Te parece razonable la estimación realizada?

$$\hat{p} = \frac{405}{500} = 0{,}81, \hat{q} = 1 - 0{,}81 = 0{,}19, n = 500$$

a) $1 - \alpha = 0{,}965 \rightarrow z_{\frac{\alpha}{2}} = 2{,}11$

$$\hat{p} \pm z_{\frac{\alpha}{2}} \cdot \sqrt{\frac{p \cdot q}{n}} = 0{,}81 \pm 2{,}11 \cdot \sqrt{\frac{0{,}81 \cdot 0{,}19}{500}} = [0{,}773; 0{,}847]_{96{,}5\ \%}$$

b) Según los datos del INE, tenemos que $\mu = 81\ \%$.

Como $\mu \epsilon [0{,}773; 0{,}847]_{96{,}5\ \%}$, entonces sí que es razonable esta estimación realizada para un nivel de significación del 96,5 %.

54. Se desea estimar la percepción media (en una escala del 0 al 10) sobre la incidencia del acoso escolar en una comunidad autónoma. Por estudios realizados anteriormente, se sabe que la desviación típica es de 3,2. Determina el tamaño muestral necesario con un nivel de confianza del 90 % y un error de ± 0,5 puntos.

Se tienen los siguientes datos:

$$\sigma = 3{,}2;\ z_{\alpha/2} = 1{,}645; E = 0{,}5$$

El tamaño muestral para estimar μ en una población con σ conocida es:

$$n = \frac{z^2_{\alpha/2} \cdot \sigma^2}{E^2} = \frac{1{,}645^2 \cdot 3{,}2 \cdot 3{,}2}{0{,}5^2} = 110{,}84$$

El tamaño muestral necesario es de 111 personas.

55. En la ficha técnica de una encuesta figura la siguiente información: «Muestra diseñada para 5400 entrevistas, a un nivel de confianza del 95,5 % (dos sigmas) y P = Q, y un error del muestreo de ± 1,6 % para el supuesto de muestreo aleatorio simple». En realidad, ¿a cuántas personas se realizó la entrevista?

- Nivel de confianza: $1 - \alpha = 0{,}955 \rightarrow z_{\frac{\alpha}{2}} = 2{,}005$

- p = q = 0,5

- Error máximo: $E = \pm 1{,}6\ \% = \pm 0{,}016$

El tamaño muestral con esta información es:

$$n = \frac{z^2_{\alpha/2} \cdot p \cdot q}{E^2} = \frac{2{,}005^2 \cdot 0{,}5 \cdot 0{,}5}{0{,}016^2} = 3925{,}8$$

En realidad se entrevistó a 3926 personas.

56. En un estudio del Centro de Investigaciones Sociológicas (CIS) aparece la siguiente información en su ficha técnica: «Tamaño de la muestra realizada a 2543 entrevistas, un nivel de confianza del 95,5 % (dos sigmas) y P = Q». Calcula el error del muestreo para el conjunto de la muestra en el supuesto de un muestreo aleatorio simple.

- Nivel de confianza: $1 - \alpha = 0{,}955 \rightarrow z_{\frac{\alpha}{2}} = 2{,}005$

- p = q = 0,5

- Tamaño muestral: n = 2543 entrevistas

El error del muestreo es:

$$E = z_{\alpha/2} \cdot \sqrt{\frac{p \cdot q}{n}} = 2{,}005 \cdot \sqrt{\frac{0{,}5 \cdot 0{,}5}{2543}} = 0{,}01987$$

Por tanto, el error del muestreo es, aproximadamente, del 2 %.

Contrastes de hipótesis

57. La presidenta de un grupo de agencias de viaje asegura que, durante el pasado mes de julio, obtuvieron unas ventas medias de 3 millones de euros. Con el fin de contrastar este dato, se tomó una muestra aleatoria de 100 sucursales y se obtuvo unas ventas medias de 2,9 millones de euros y una desviación típica de 0,35 millones de euros. ¿Se puede aceptar la afirmación de la presidenta con un nivel de confianza del 95 %?

Población

Muestra: $\begin{cases} n = 100 \\ \bar{x} = 2{,}9\ millones\ de\ € \\ \hat{s} = 0{,}35\ millones\ de\ € \end{cases}$

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

Construimos un intervalo de confianza para la media poblacional μ:

$$\mu \epsilon \left[\bar{x} - z_{\frac{\alpha}{2}} \cdot \frac{\hat{s}}{\sqrt{n}}, \bar{x} + z_{\frac{\alpha}{2}} \cdot \frac{\hat{s}}{\sqrt{n}}\right] = [2{,}8314; 2{,}9686]_{95\ \%}$$

Como $\mu \notin [2{,}8314; 2{,}9686]_{95\ \%}$, entonces no se acepta la afirmación de la presidenta para un nivel de confianza del 95 %.

58. Se ha estudiado el ahorro anual (deuda en el caso de valores negativos) de las familias de una determinada comunidad autónoma, y se ha modelado por una distribución normal con 20 000 € de desviación típica.

a) Se elige una muestra aleatoria de 25 familias y se obtiene un ahorro medio anual de 5000 €. Determina sendos intervalos de confianza del 90 y 95 % para el ahorro medio anual familiar de esta comunidad.

b) Teniendo en cuenta los resultados del apartado anterior, ¿se podría afirmar que las familias de esta comunidad no ahorran?

c) Si se desea obtener un intervalo de confianza al 90 % para el ahorro medio familiar con una amplitud de 20 000 €, ¿cuántas personas habría que seleccionar en la muestra?

a) Población, N(μ, σ): $\begin{cases} \mu = ? \\ \sigma = 20\,000\ € \end{cases}$

Muestra: $\begin{cases} n = 25 \\ \bar{x} = 5000\ € \end{cases}$

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

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

El intervalo de confianza para la altura media μ al 90 % es:

$$\left[\bar{x} - z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}, \bar{x} + z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}\right] = [-1580; 11\,580]_{90\,\%}$$

El intervalo de confianza para la altura media μ al 95 % es:

$$\left[\bar{x} - z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}, \bar{x} + z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}\right] = [-2840; 12\,840]_{95\,\%}$$

b) No se puede descartar que las familias no ahorren, ya que en ambos intervalos de confianza hay valores negativos. Por tanto, no se rechaza la afirmación de que las personas de esta comunidad autónoma no ahorran.

c) $L = 2 \cdot z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}} \to 20\,000 = 2 \cdot 1{,}645 \cdot \frac{20\,000}{\sqrt{n}} \to n = 10{,}82$

Habría que seleccionar en la muestra a 11 familias.

Aplicaciones

59. Baterías de automóviles. Un fabricante de automóviles afirma que la duración de sus baterías sigue una distribución normal de media 85 000 km y una desviación típica de 6000 km. En una muestra aleatoria de 25 baterías se obtuvo una duración media de 83 200 km. Suponiendo que la afirmación del fabricante es correcta, ¿cuál es la probabilidad de obtener una media muestral igual o inferior a la de la muestra?

Población, N(μ, σ): $\begin{cases} \mu = 85\,000\ km \\ \sigma = 6000\ km \end{cases}$

Muestra: $\begin{cases} n = 25 \\ \bar{x} = 83\,200\ km \end{cases}$

La distribución muestral de la media μ es:

$$\mu_{\bar{x}} = \mu = 85\,000;\ \sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{6000}{5} = 1200$$

$\bar{x} \sim N(85\,000; 1200)$, o bien: $z = \frac{\bar{x} - 85\,000}{1200} \sim N(0{,}1)$

$$p(\bar{x} \le 83\,200) = p\left(z \le \frac{83\,200 - 85\,000}{1200}\right) = p(z \le -1{,}25) = 0{,}1056\ o\ 10{,}56\,\%$$

60. **Taladros eléctricos.** **Una cadena de ferreterías ha recibido en un contenedor 5000 taladros eléctricos que el fabricante indica que se distribuyen según una normal de media 295 W y una desviación típica de 12 W. Antes de aceptar el envío, el encargado selecciona una muestra al azar de 9 taladros para probarlos y analizar el consumo máximo de energía de cada uno, de tal forma que, si el consumo medio de la muestra es superior a 300 W (valor que figura en la etiqueta del producto), rechazará el envío. ¿Cuál es la probabilidad de que se rechace el contenedor de taladros?**

Población, N(μ,σ): $\begin{cases}\mu = 295\ kW \\ \sigma = 12\ kW\end{cases}$

Muestra: $n = 9$

La distribución de la media muestral $\bar{x}$ es:

$$\mu_{\bar{x}} = \mu = 295;\ \sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{12}{3} = 4$$

$\bar{x} \sim N(295; 4)$, o bien: $z = \frac{\bar{x}-295}{4} \sim N(0{,}1)$

$$p(\bar{x} > 300) = p\left(z > \frac{300-295}{4}\right) = p(z > 1{,}25) = 0{,}1056\ o\ 10{,}56\ \%$$

La probabilidad de que se rechace el contenedor de taladros es del 10,56 %.

61. **Envases de azúcar.** **Se dispone de una máquina para llenar envases de azúcar. La cantidad de azúcar depositada en el envase es una variable aleatoria normal de media 500 g y una desviación típica de 15 g. Para verificar que el peso medio se mantiene en 500 g, se toman muestras de 25 cajas y se pesa su contenido. Se decide detener el proceso de envasado cada vez que el peso medio de la muestra sea inferior a 495 g o superior a 505 g. ¿Cuál es la probabilidad de que se detenga el proceso?**

Población, N(μ,σ): $\begin{cases}\mu = 500\ g \\ \sigma = 15\ g\end{cases}$

Muestra: $n = 25\ cajas$

La distribución de la media muestral $\bar{x}$ es:

$$\mu_{\bar{x}} = \mu = 500;\ \sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{15}{5} = 3$$

$\bar{x} \sim N(500; 3)$, o bien: $z = \frac{\bar{x}-500}{3} \sim N(0{,}1)$

$$p(495 \leq \bar{x} \leq 505) = p\left(\frac{495-500}{3} \leq z \leq \frac{505-500}{3}\right) = p(-16{,}7 \leq z \leq 1{,}67)$$
$$= 0{,}9050 \; o \; 90{,}50\ \%$$

Por tanto, la probabilidad de que se detenga el proceso es:

1 – 0,9050 = 0,0950

En otras palabras: en el 95 % de los casos se detendrá el proceso de envasado.

62. Instalaciones renovables. **Una empresa de instalaciones renovables está estudiando la localización para instalar energía eólica. Por experiencia en otras zonas similares, la empresa sabe que la energía eólica diaria sigue una distribución normal. En una zona similar se obtuvo en una muestra de 36 días una media de 42 kWh y una desviación típica de 9 kWh. Construye un intervalo de confianza al 95 % para estimar la energía eólica diaria de la nueva localización.**

Población: N(μ, σ)

Muestra: $\begin{cases} \mu = 3{,}6\ días \\ \bar{x} = 42\ kWh \\ \hat{s} = 9\ kWh \end{cases}$

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

El intervalo de confianza para la media poblacional μ de una población normal con σ desconocida es:

$$\left[\bar{x} - z_{\frac{\alpha}{2}} \cdot \frac{\hat{s}}{\sqrt{n}}, \bar{x} + z_{\frac{\alpha}{2}} \cdot \frac{\hat{s}}{\sqrt{n}}\right] = [39{,}06; 44{,}94]_{95\ \%}$$

63. Control de calidad. **Por información del control de calidad de una empresa se sabe que el peso de los paquetes envasados por una máquina sigue una distribución normal con una desviación típica de 0,63 g. Se desea seleccionar una muestra aleatoria simple de paquetes envasados por esa máquina con el fin de determinar el peso medio, con una confianza del 95 % y con la condición de que la media muestral no difiera de la media poblacional en más de 0,1 g. ¿Qué tamaño de muestra se necesita?**

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

Población: N(μ, σ): $\sigma = 0{,}63\ g$

Muestra: $\begin{cases} \mu = ? \\ \bar{x} = ? \end{cases}$

La distribución muestral de la media $\bar{x}$ viene dada por:

$$\mu_{\bar{x}} = \mu; \ \sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{0{,}63}{\sqrt{n}}$$

$\bar{x} \sim N(\mu; \frac{0{,}63}{\sqrt{n}})$, o bien: $z = \frac{\bar{x}-\mu}{\frac{0{,}63}{\sqrt{n}}} \sim N(0{,}1)$

El intervalo de confianza para μ es:

$$\mu = \bar{x} \pm 1{,}96 \cdot \frac{0{,}63}{\sqrt{n}} \rightarrow |\mu - \bar{x}| = \pm 1{,}96 \cdot \frac{0{,}63}{\sqrt{n}}$$

Como $|\mu - \bar{x}| \leq 0{,}1$, entonces:

$$1{,}96 \cdot \frac{0{,}63}{\sqrt{n}} \leq 0{,}1 \rightarrow n \geq \left(\frac{1{,}96 \cdot 0{,}63}{0{,}1}\right)^2 = 152{,}47$$

El tamaño muestral necesario tiene que ser de, al menos, 153 paquetes.

64. Aplicación móvil basada en IA. Una empresa va a lanzar al mercado una nueva aplicación móvil, basada en inteligencia artificial, para cuidar la salud de las mascotas. Las aplicaciones lanzadas por esta empresa tienen una valoración que sigue una distribución normal. Con el fin de medir la aceptación de esta nueva aplicación, realiza una encuesta a 10 personas que ya la han probado y dan las siguientes valoraciones (1-10):

8,5	9	8	7	8	9	6	7	7,5	4

a) Da una estimación puntual para la valoración media y la desviación típica.

b) Da una estimación para la valoración media mediante un intervalo de confianza al 99 %.

a) $\bar{x} = \frac{\sum_{i=1}^{10} x_i}{10} = \frac{74}{10} = 7{,}4$

Y una estimación puntual de la desviación típica es la cuasidesviación típica o desviación típica muestral. Así pues:

$\hat{s} = \sqrt{\frac{\sum_{i=1}^{10}(x_i-\bar{x})^2}{n-1}} = 1{,}52$

b) Población: N(μ, σ): σ

Muestra: $\begin{cases} n = 10 \\ \bar{x} = 7{,}4 \\ \hat{s} = 1{,}52 \end{cases}$

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

El intervalo de confianza para la media poblacional μ es:

$$\bar{x} \pm z_{\frac{\alpha}{2}} \cdot \frac{\hat{s}}{\sqrt{n}} = [6{,}2; 8{,}6]_{99\,\%}$$

65. **Ofertas de empleo. Una consultoría se dedica a ofertar empleos y encontró que, en una muestra aleatoria de 60 empresas, el 30 % consulta los perfiles de los posibles candidatos en la red social LinkedIn antes de contratarlos. Construye un intervalo de confianza para la proporción poblacional de las empresas que realizan esa consulta, a un nivel de confianza del 90 %.**

$$\hat{p} = 0{,}3, \hat{q} = 1 - 0{,}3 = 0{,}7, n = 60$$

$$1 - \alpha = 0{,}90 \rightarrow z_{\frac{\alpha}{2}} = 1{,}645$$

El intervalo de confianza para la proporción poblacional p es:

$$\hat{p} \pm z_{\frac{\alpha}{2}} \cdot \sqrt{\frac{\hat{p} \cdot \hat{q}}{n}} = 0{,}3 \pm 1{,}645 \cdot \sqrt{\frac{0{,}3 \cdot 0{,}7}{60}} = [0{,}2027; 0{,}3973]_{90\,\%}$$

Con un nivel de confianza del 90 %, podemos indicar que las empresas consulta en el perfil de los candidatos en la red social LinkedIn entre el 20,27 y el 39,73 % de las ocasiones.

66. **Audiencia TV. Para realizar un estudio sobre la audiencia de una cadena de televisión local, se han obtenido los siguientes resultados sobre el número de residentes en las diferentes zonas:**

Zonas	Residentes
A	15 320
B	32 416
C	12 140
D	6 734
Total	66 610

a) Distribuye una muestra de 400 entrevistas para hacer este estudio, mediante un diseño estratificado de afijación proporcional al tamaño de cada zona.

b) Para la zona C, calcula un intervalo de confianza al 95 % para la proporción de televidentes de esta cadena local, utilizando el resultado muestral hallado en a), sabiendo que el 35 % de los entrevistados ven con asiduidad ese canal.

a) Sean n_1, n_2, n_3 y n_4 el número de entrevistas en las zonas A, B, C y D, respectivamente.

$$\frac{n}{N}=\frac{n_1}{N_A}=\frac{n_2}{N_B}=\frac{n_3}{N_C}=\frac{n_4}{N_D}\to\frac{400}{66\,610}=\frac{n_1}{15\,320}=\frac{n_2}{32\,416}=\frac{n_3}{12\,140}=\frac{n_4}{6734}$$

Luego: $n_1=92$, $n_2=195$, $n_3=73$ y $n_4=40$.

b) $\hat{p}=0{,}35, \hat{q}=1-0{,}35=0{,}65, n_3=73$

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

El intervalo de confianza para la proporción p de televidentes en la zona C es:

$$\hat{p}\pm z_{\frac{\alpha}{2}}\cdot\sqrt{\frac{\hat{p}\cdot\hat{q}}{n_3}}=0{,}35\pm 1{,}96\cdot\sqrt{\frac{0{,}35\cdot 0{,}65}{73}}=[0{,}2406;0{,}4594]_{95\,\%}$$

67. Inflación. Con el fin de analizar la preocupación por la inflación de los ciudadanos de la Unión Europea, se seleccionó una muestra aleatoria de 2100 personas, de las cuales 1512 manifestaron que se sentían bastante o muy preocupadas por la inflación. Encuentra un intervalo de confianza al 95 % para la proporción de ciudadanos de la Unión Europea que se sienten bastante o muy preocupados por la inflación.

$$\hat{p}=0{,}72, \hat{q}=1-0{,}72=0{,}18, n=2100\,personas$$

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

El intervalo de confianza para la proporción poblacional p es:

$$\hat{p}\pm z_{\frac{\alpha}{2}}\cdot\sqrt{\frac{\hat{p}\cdot\hat{q}}{n}}=0{,}72\pm 1{,}96\cdot\sqrt{\frac{0{,}72\cdot 0{,}18}{2100}}=[0{,}7008;0{,}7392]_{95\,\%}$$

El porcentaje de ciudadanos de la UE que se sienten bastante o muy preocupados por la inflación se encuentra entre el 70,08 y el 73,92 %, con un nivel de confianza del 95 %.

68. Alcohol y patinetes. Para informar y concienciar a los jóvenes sobre lo nocivo que es el alcohol, la policía local de un municipio efectúa controles de alcoholemia a los jóvenes que circulan en patinete por la noche. Para ello, decide elegir aleatoriamente una muestra de 50 jóvenes, de los cuales 10 superan la tasa del alcohol permitida.

a) Construye un intervalo de confianza para la proporción de jóvenes que utilizan patinetes y superan la tasa de alcohol permitida con un nivel de confianza del 90 %.

b) ¿Cuál debería ser el tamaño muestral, manteniendo el mismo nivel de confianza, para duplicar la precisión de la estimación realizada?

$$\hat{p} = 0{,}2, \hat{q} = 1 - 0{,}2 = 0{,}8, n = 50$$

$$1 - \alpha = 0{,}90 \rightarrow z_{\frac{\alpha}{2}} = 1{,}645$$

El intervalo de confianza para la proporción poblacional p es:

$$\hat{p} \pm z_{\frac{\alpha}{2}} \cdot \sqrt{\frac{\hat{p} \cdot \hat{q}}{n}} = 0{,}2 \pm 1{,}645 \cdot \sqrt{\frac{0{,}2 \cdot 0{,}8}{50}} = [0{,}1069; 0{,}2931]_{90\,\%}$$

b) El error mínimo de la estimación en el caso anterior es:

$$E = z_{\frac{\alpha}{2}} \cdot \sqrt{\frac{\widehat{p} \cdot \widehat{q}}{n}} = 1{,}645 \sqrt{\frac{0{,}2 \cdot 0{,}8}{50}} = 0{,}0931$$

Si se desea duplicar la precisión, entonces el error debe ser la mitad, es decir, E = 0,04655.

Por tanto, el tamaño muestral requerido manteniendo el mismo nivel de confianza y considerando p = q =0,5 es:

$$n = \frac{z_{\alpha/2}^2 \cdot \widehat{p} \cdot \widehat{q}}{E'^2} = 312{,}19 \approx 312$$

El tamaño muestral debería ser de 312 jóvenes, es decir, algo más de seis veces más que en el caso anterior.

69. Sostenibilidad. Las emisiones diarias específicas de CO_2, medidas en kg/MWh, de las centrales de carbón de una empresa generadora de electricidad se pueden modelizar por una distribución normal. Al elegir una muestra de 45 días en una de sus centrales de carbón, se obtuvo una media de emisiones de 766,5 kg/MWh y una desviación típica de 23,4 kg/MWh.

a) Construye un intervalo de confianza al 95 % para la media diaria de las emisiones específicas.

b) La empresa indica que las emisiones medias diarias de sus centrales son de 785 kg/MWh. ¿Se puede aceptar esta hipótesis al 95 %?

Población: N(μ, σ)

Muestra: $\begin{cases} n = 45\ días \\ \bar{x} = 766{,}5\ kg/MWh \\ \hat{s} = 23{,}4\ kg/MWh \end{cases}$

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

El intervalo de confianza para la media poblacional μ de una población con σ desconocida es:

$$\bar{x} \pm z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}} = [759{,}66; 773{,}34]_{95\,\%}$$

La media de emisiones de CO_2 estará comprendida entre los dos extremos de ese intervalo.

b) Como $\mu = 785 \notin [759{,}66; 773{,}34]_{95\,\%}$, se rechaza la hipótesis planteada, para un nivel de confianza del 95%.

Prepárate para la universidad

70. **El consumo de energía eléctrica mensual por vivienda medido en kilovatios hora (kWh) sigue una distribución normal con varianza 4225 $(kWh)^2$.**

a) Se toma una muestra aleatoria de 100 viviendas, obteniéndose un consumo total de 26 830 kWh. Calcula un intervalo de confianza al 92 % para estimar el consumo medio poblacional.

b) Calcula el tamaño mínimo de la muestra necesario para estimar el consumo medio de energía eléctrica mensual por vivienda, con un error máximo de 5 kWh y un nivel de confianza del 98 %.

c) Tras una campaña para incentivar el ahorro energético, se toma una nueva muestra y el intervalo de confianza para el consumo medio que se obtiene es (224,08 ; 255,92). Calcula la media del consumo de energía eléctrica mensual por vivienda para dicha muestra.

Andalucía

$a)\ Población: \begin{cases} \mu = ? \\ \sigma = 65\ kWh \end{cases},$ N(μ, σ)

Muestra: $\begin{cases} n = 100 \\ \bar{x} = \frac{26\,830}{100} = 268{,}3\ kWh \end{cases}$

$$1 - \alpha = 0{,}92 \rightarrow z_{\frac{\alpha}{2}} = 1{,}75$$

El intervalo de confianza al 92 % para estimar el consumo medio poblacional μ es:

$$\bar{x} \pm z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}} = 268{,}3 \pm 1{,}75 \cdot \frac{65}{10} = 268{,}3 \pm 11{,}375 = [256{,}925; 279{,}675]_{92\,\%}$$

El consumo de energía eléctrica mensual por vivienda variará entre esos valores del intervalo para un nivel de confianza del 92 %.

b) $1 - \alpha = 0{,}98 \rightarrow \alpha = 0{,}02 \rightarrow z_{\frac{\alpha}{2}} = 2{,}33$

El tamaño muestral, n, viene dado por:

$$n = \left(\frac{z_{\alpha/2} \cdot \sigma}{E}\right)^2 = \left(\frac{2{,}33 \cdot 65}{5}\right)^2 = 917{,}48$$

El tamaño mínimo de la muestra debe ser de 918 hogares.

c) El intervalo facilitado para μ en este caso es:

[224,08; 255,92]

Por tanto:

$$\bar{x} = \frac{224{,}08 + 255{,}92}{2} = 240\ kWh$$

La media del consumo de energía eléctrica mensual por vivienda en dicha muestra es de 240 kWh.

71. **Se quiere hacer un estudio para estimar la proporción de personas que ha viajado a América.**

a) ¿Cuál será el tamaño muestral necesario para que pueda estimarse la verdadera proporción de personas que han viajado a América, a partir de la proporción muestral, con un error de estimación máximo de 0,05 y un nivel de confianza del 90 %?

b) En una muestra aleatoria de 2000 personas, se sabe que 600 han viajado a América. En función de esta muestra obtiene, con un nivel de confianza del 90 %, un intervalo para estimar la proporción poblacional de personas que han viajado a América.

Aragón

a) $Muestra: \begin{cases} \hat{p}\ ? \\ n\ ? \\ 1 - \alpha = 0{,}90 \to z_{\frac{\alpha}{2}} = 1{,}645 \end{cases}$

Como no se puede estimar $\hat{p}$, se considera el caso más desfavorable, que corresponde a $\hat{p} = \hat{q} = 0{,}5$. Así, el tamaño muestral sería:

$$n = \left(\frac{z_{\alpha/2}}{E}\right)^2 \cdot (\hat{p} \cdot \hat{q}) = \left(\frac{1{,}645}{0{,}05}\right)^2 \cdot (0{,}5 \cdot 0{,}5) = 270{,}60$$

La muestra debería ser de 271 personas.

b) $Muestra: \begin{cases} n = 2000 \\ \hat{p} = \frac{600}{2000} = 0{,}3;\ \hat{q} = 0{,}7 \end{cases}$

$$1 - \alpha = 0{,}90 \to z_{\frac{\alpha}{2}} = 1{,}645$$

El intervalo para estimar la proporción poblacional, p, viene dado por:

$$\hat{p} \pm z_{\frac{\alpha}{2}} \cdot \sqrt{\frac{\hat{p} \cdot \hat{q}}{n}} = 0{,}3 \pm 1{,}645 \cdot \sqrt{\frac{0{,}3 \cdot 0{,}7}{2000}} = 0{,}3 \pm 0{,}0169 = [0{,}2831; 0{,}3169]_{90\,\%}$$

Con una confianza del 90 %, el porcentaje de personas que viajó a América está comprendido entre el 28,31 % y el 31,69 %.

72. **Se sabe que las pruebas de acceso a la Universidad realizadas por los estudiantes de segundo de Bachillerato siguen una distribución normal.**

a) Si la media de la prueba selectiva es de 65 puntos y la desviación típica es de 8, calcula la probabilidad de que la nota media de 25 estudiantes elegidos al azar sea mayor de 63 puntos.

b) Calcula un intervalo de confianza para la nota media de ingreso en Doble Grado en Derecho y ADE, con un nivel de confianza del 92 %, sabiendo que ingresan 100 estudiantes, que la nota media de acceso es de 80 puntos y que la desviación típica es de 8,8 puntos.

c) Determina el tamaño de la muestra necesario para que el error máximo del intervalo de confianza calculado en el apartado anterior se reduzca a la mitad (con los datos de b)).

Asturias

a) $Población: \begin{cases} \mu = 65 \\ \sigma = 8 \end{cases}$, N$(\mu, \sigma)$

Muestra: $n = 65$

La distribución muestral de la media $\bar{x}$ es:

$$\bar{x} \sim N\left(\mu, \frac{\sigma}{\sqrt{n}}\right) = \left(65, \frac{8}{\sqrt{65}}\right) = (65;\ 1{,}6)$$

O sea:

$$z = \frac{\bar{x} - 65}{1{,}6} \sim N(0{,}1)$$

$$p(\bar{x} > 63) = p\left(\frac{\bar{x} - 6{,}5}{1{,}5} > \frac{63 - 65}{1{,}6}\right) = p(z > -1{,}25) = p(z < 1{,}25) = 0{,}8944\ o\ 89{,}44\ \%$$

La probabilidad de que la nota media de 25 estudiantes elegidos al azar sea superior a 63 puntos es del 89,44 %.

b) $Población: \begin{cases} \mu = ? \\ \sigma = ? \end{cases}$, N$(\mu, \sigma)$

$$Muestra: \begin{cases} n = 100 \\ \bar{x} = 80 \\ \hat{s} = 8{,}8 \end{cases} \rightarrow z_{\frac{\alpha}{2}} = 1{,}75$$

Así pues:

$$\left[\bar{x} - z_{\frac{\alpha}{2}} \cdot \frac{\hat{s}}{\sqrt{n}}, \bar{x} + z_{\frac{\alpha}{2}} \cdot \frac{\hat{s}}{\sqrt{n}}\right] = \left[80 - 1{,}75 \cdot \frac{8{,}8}{10}, 80 + 1{,}75 \cdot \frac{8{,}8}{10}\right] = [78{,}46; 81{,}54]_{92\ \%}$$

A un nivel de confianza del 92 %, se estima que la nota media de acceso estará entre 78,46 y 81,54 puntos.

c) El error máximo de la solución del apartado anterior es:

$$E = z_{\frac{\alpha}{2}} \cdot \frac{\hat{s}}{\sqrt{n}} = 1{,}75 \cdot \frac{8{,}8}{10} = 1{,}54$$

Por tanto, el nuevo error debe ser:

$$E' =\cdot \frac{1{,}54}{2} = 0{,}77$$

De este modo, el tamaño muestral vale:

$$n = \left(\frac{z_{\alpha/2} \cdot \hat{s}}{E'}\right)^2 = \left(\frac{1{,}75 \cdot 0{,}88}{0{,}77}\right)^2 = 400$$

El tamaño muestral mínimo debe ser de 400 estudiantes para que el error máximo sea de 0,77 puntos.

73. La enóloga de una bodega ha determinado que el porcentaje de alcohol presente en sus botellas de vino sigue una distribución normal con una desviación típica de 0,53 %. Una muestra de 120 botellas, escogidas al azar, arroja un valor promedio para el porcentaje de alcohol por botella de 12,05 %.

a) Obtén el intervalo de confianza del 95 % para el valor promedio del porcentaje de alcohol por botella.

b) ¿Cuál es el número mínimo de botellas que habría que considerar para que el error cometido al estimar el valor medio del porcentaje de alcohol por botella, con un nivel de confianza del 97,5 %, fuera del 0,1 %?

Cantabria

$Población$: $\begin{cases} \mu = ? \\ \sigma = 0{,}53 \end{cases}$, N$(\mu, \sigma)$

$$Muestra: \begin{cases} n = 120\ botellas \\ \bar{x} = 12{,}05\ \% \end{cases}$$

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

El intervalo de confianza para la media $\mu\ es$:

$$\left[\bar{x} - z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}, \bar{x} + z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}\right] = 12{,}05 \pm 1{,}96 \cdot \frac{0{,}53}{\sqrt{120}} = 12{,}05 \pm 0{,}09 = [11{,}96; 12{,}14]_{95\ \%}$$

El porcentaje medio de alcohol por botella oscila entre el 11,96 % y el 12,14 % para un nivel de confianza del 95 %.

b) Si $1 - \alpha = 0{,}975 \rightarrow \alpha = 0{,}025 \rightarrow z_{\frac{\alpha}{2}} = 2{,}24$

Por tanto, el tamaño muestral, n, vendrá dado por:

$$n = \left(\frac{z_{\alpha/2} \cdot \sigma}{E'}\right)^2 = \left(\frac{2{,}24 \cdot 0{,}53}{0{,}1}\right)^2 = 140{,}94$$

La muestra debería tener, al menos, 141 botellas.

74. El número de pacientes que se atienden semanalmente en un centro de salud sigue una distribución normal de media desconocida y una desviación típica de σ = 50 pacientes. Se ha tomado una muestra aleatoria de 25 semanas y se ha registrado el número de pacientes atendidos, proporcionando una media de 322 pacientes.

a) Calcula un intervalo de confianza para la media poblacional del número de pacientes atendidos con un nivel de confianza del 95 %.

b) Explica razonadamente qué ocurrirá con la amplitud del intervalo si, para el mismo nivel de confianza, aumentamos el tamaño muestral.

c) ¿Se puede aceptar la afirmación de que la media de pacientes atendidos a la semana es de 330 con un nivel de confianza del 99 %? Justifica la respuesta.

Castilla-La Mancha

$$Población: \begin{cases} \mu = ? \\ \sigma = 50\ pacientes \end{cases}, \text{N}(\mu, \sigma)$$

$$Muestra: \begin{cases} n = 25 \\ \bar{x} = 322\ pacientes \end{cases}$$

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

El intervalo de confianza para la media poblacional $\mu\ es$:

$$\left[\bar{x} - z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}, \bar{x} + z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}\right] = \left[322 - 1{,}96 \cdot \frac{50}{5}; 322 + 1{,}96 \cdot \frac{50}{5}\right] = [302{,}4; 341{,}6]_{95\,\%}$$

El número medio de pacientes atendidos está comprendido entre esos dos valores límite para un nivel de confianza del 95 %.

b) Si se aumenta el tamaño muestral, el error muestral disminuirá y, por tanto, la amplitud del intervalo también lo hará, luego ganaremos en precisión.

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

El intervalo de confianza para la media poblacional $\mu\ es$:

$$\mu \epsilon \left[\bar{x} - z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}, \bar{x} + z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}\right] = [322 - 25{,}75; 322 + 25{,}75] = [296{,}25; 347{,}75]_{99\,\%}$$

Como $\mu = 300 \epsilon [296{,}25; 347{,}75]_{99\,\%}$, no se rechaza la afirmación de que la media de pacientes atendidos a la semana es de 330 con un nivel de confianza del 99 %.

75. La ficha técnica de una encuesta electoral realizada para las pasadas elecciones autonómicas indica que se ha encuestado a 1000 individuos con derecho a voto residentes en Castilla y León. La muestra se ha tomado mediante un muestreo aleatorio simple. El error de la estimación de la proporción de individuos de la población que vota al partido K es de ± 3,2 %, fijada una confianza del 95,5 %. Para esta ficha técnica, identifica los siguientes elementos: población, diseño muestral, tamaño muestra y parámetro estimado.

Castilla y León

Población: residentes en Castilla y León en derecho a voto.

Muestra: 1000 individuos

Diseño muestral: muestreo aleatorio simple

Nivel de confianza:

$$1 - \alpha = 0{,}955 \rightarrow \alpha = 0{,}045 \rightarrow z_{\frac{\alpha}{2}} = 2{,}005$$

Parámetro estimado: población p que vota al partido K.

Aunque no lo pide el problema, vamos a comprobar que efectivamente el error máximo de la estimación es:

E = 0,032, para n = 1000, $z_{\frac{\alpha}{2}} = 2{,}005$ y, en el caso más desfavorable, donde $\hat{p} = \hat{q} = 0{,}5$. Así pues:

$$E = z_{\alpha/2} \cdot \sqrt{\frac{\hat{p} \cdot \hat{q}}{n}} = 2{,}005 \cdot \sqrt{\frac{0{,}5 \cdot 0{,}5}{1000}} = 0{,}0317 \approx 0{,}032 \text{ } o \text{ } 3{,}2\,\%$$

76. Una cadena de supermercados tiene en plantilla a 3000 cajeros, 4000 reponedores y 1000 transportistas. Se desea obtener una muestra de 200 trabajadores para una encuesta sobre satisfacción con el puesto de trabajo. Se pide, razonadamente, las respuestas:

a) Atendiendo a razones de proporcionalidad, ¿cuántos cajeros, reponedores y transportistas debería seleccionar la empresa para la encuesta?

b) Si 30 de los cajeros encuestados estaba satisfecho en su trabajo, da una estimación de la proporción de cajeros satisfechos en su puesto de trabajo.

Extremadura

a) Se trata de un muestreo aleatorio estratificado (en el que los estratos son los cajeros, los reponedores y los transportistas) con afijación proporcional. Por tanto:

$$N = N_1 + N_2 + N_3 = 3000 + 4000 + 1000 = 8000$$

$$n = 200 = n_1 + n_2 + n_3$$

$$\frac{N}{n} = \frac{N_1}{n_1} = \frac{N_2}{n_2} = \frac{N_3}{n_3} \to \frac{8000}{200} = \frac{3000}{n_1} = \frac{4000}{n_2} = \frac{1000}{n_3}$$

Es decir:

$$n_1 = 75, n_2 = 100, n_3 = 25$$

Se deberán seleccionar 75 cajeros, 100 reponedores y 25 transportistas para realizar dicha encuesta.

b) Muestra de cajeros: $\begin{cases} n_1 = 75 \\ \hat{p} = \frac{30}{75} = 0{,}4;\ \hat{q} = 0{,}6 \end{cases}$

Un 40 % de los cajeros está satifecho con su puesto de trabajo.

