Distribución de las medias muestrales. Estimación de la media

1. Se supone que la estatura de los chicos de 18 años de cierta población sigue una distribución normal de media 162 cm y desviación típica 12 cm. Se toma una muestra al azar de 100 de estos chicos encuestados y se calcula la media. ¿Cuál es la probabilidad de que esta media esté entre 159 y 165 cm? ***Solución: 0'9876***

2. Un grupo de 144 alumnos de Secundaria seleccionados al azar en una determinada comunidad realiza una prueba de conocimientos sobre la geografía de su autonomía, sacando una media de 6'3 puntos. Las puntuaciones se distribuyen normalmente con una desviación típica de 6. Calcula con una probabilidad del 98 %, entre qué valores se encontrará la media de la población de los alumnos de Secundaria de dicha comunidad. Interpreta el significado del intervalo obtenido. ***Solución:*** $\boldsymbol{z_{\frac{\alpha}{2}} = 2'33, Intervalo = (5'135, 7'465)}$

3. La longitud de la ballena azul se distribuye según una ley normal con desviación típica de 7'5 m. En un estudio estadístico realizado a 25 ejemplares se ha obtenido el intervalo de confianza $(\mathbf{21'06}, \mathbf{26'94})$ para la longitud media. **(a)** Calcula la longitud media de los 25 ejemplares de la muestra. **(b)** Calcula el nivel de confianza el que se ha construido dicho intervalo. ***Solución: (a)*** $\boldsymbol{\overline{x} = 24}$ ***(b) El nivel de confianza es del 95 %.***

4. Con el fin de estimar la edad media de los habitantes de una gran ciudad, se tomó una muestra aleatoria de 300 habitantes de una gran ciudad, que arrojó una edad media de 35 años y una desviación típica de 7 años. **(a)** Halla el intervalo del 95 % de confianza en el que se encontrará la edad media de la población. **(b)** Para que el intervalo fuera $35 \pm 0'44$, ¿qué nivel de confianza debería usar? ***Solución: (a) (34'20, 35'9) (b) 72?42 %***

5. En un estudio sobre la longevidad de los habitantes de una comunidad se contabilizan 121 personas para las que se obtiene una media de 79'5 años de vida. **(a)** Si se maneja una desviación típica igual a 3'5 años y un nivel de significación del 3 %, construye el intervalo de confianza para la longevidad media de los habitantes de la comunidad. **(b)** Con la misma desviación típica del apartado anterior y con un nivel de confianza del 99 %, ¿cuál debería ser el tamaño de la muestra para que la amplitud del intervalo de confianza fuera igual a un año? ***Solución: (a)*****(78'81, 80'19)** ***(b)n = 325.***

6. Se sabe que los estudiantes de una provincia duermen un número de horas diarias que se distribuye según una ley normal de media m horas y desviación típica 2 horas. **(a)** A partir de una muestra de 64 alumnos se ha obtenido el siguiente intervalo de confianza (7'26, 8'14) para la media de la población. Determina el nivel de confianza con que se ha construido el intervalo. **(b)** Determina el tamaño muestral mínimo necesario para que el error que se cometa al estimar la media de la población por un intervalo de confianza se, como máximo, de 0'75 horas, con un nivel de confianza del 98 %. ***Solución: (a) 92'16 % (b) n = 39.***

Distribución de las proporciones muestrales. Estimación de una proporción

7. En una determinada población se toma una muestra al azar de 256 personas. De esta muestra, el 20 % de las personas lleva gafas graduadas y el 80 % restante, no. Calcula el intervalo de confianza aproximado para la proporción poblacional de las personas que llevan gafas graduadas para un nivel de confianza del 95 %. ***Solución: (0'151, 0'249). Entre el 15'1 % y el 24'9 %.***

8. En una universidad se toma al azar una muestra de 100 alumnos y se encuentra que han suspendido todas las asignaturas 10 alumnos. Halla con un nivel de confianza del 95 %, un intervalo para estimar el porcentaje de alumnos que aprueba, al menos, una asignatura. ***Solución: (0'8412, 0'9588).***

9. En una muestra de 600 personas de una ciudad se observa que 30 son inmigrantes. **(a)** Determina un intervalo de confianza de nivel 0'95 para el porcentaje de inmigrantes en la ciudad. **(b)** Si se quiere estimar el porcentaje de inmigrantes con un error máximo de 0'02, ¿cuál es el tamaño de la muestra que habría que considerar si se usa un nivel de significación del 1 %? ***Solución: (a) (0'032, 0'068) Entre el 3'3 % y el 6'8 % (b) n = 788***

10. En una población de 3000 alumnos de Bachillerato se quiere estimar la proporción de alumnos favorables al examen de selectividad. Se selecciona, para ello, a través de muestreo estratificado aleatorio con afijación proporcional, una muestra de 600 alumnos de dicha población. Los estratos son las cuatro modalidades de bachillerato (Artes (A), Ciencias de la Naturaleza y de la Salud (B), Humanidades y Ciencias Sociales (C) y Tecnología (D)). Sabiendo que en la población hay 300 alumnos de A, 1200 de B y 400 de D, y que recogida la información un total de 450 alumnos se han mostrado favorables a la selectividad, determina: **(a)** El número de alumnos seleccionados en cada estrato. **(b)** La proporción estimada de alumnos en esa población que no son favorables a la selectividad. **(c)** El error máximo cometido en la estimación anterior con una confianza del 95 %. ***Solución: (a) EA=60, EB=240, EC=220, ED=80 (b)*** $\frac{150}{600}=\frac{1}{4}$ ***(c) Error=0'0346***

Contrastes de hipótesis

11. Las tensiones de ruptura de los cables fabricados por una empresa, tiene media de 1800 Nw y una desviación típica de 100 Nw. Se desea comprobar si un nuevo proceso de fabricación modifica dicha tensión media de ruptura. Para ello, se toma una muestra de 50 cables y se encuentra que su tensión media de ruptura es 1850 Nw. ¿Se puede afirmar que el nuevo proceso ha modificado la tensión media de ruptura al nivel de significación del 5 %? ***Solución: Como*** $\mathbf{1850} \notin (\mathbf{1772'3}, \mathbf{1827'7})$ ***debemos rechazar la hipótesis nula y aceptar la alternativa. Por tanto, el nuevo proceso sí ha modificado la tensión media de ruptura.***

12. Según un estudio realizado por una empresa hotelera durante el año 2011, la distribución del tiempo de estancia de cada viajero fue normal con una media de 3'7 días y una desviación típica de 1'1 días. A lo largo del año 2012 se analizó el tiempo de estancia de 49 viajeros elegidos al azar, obteniéndose una media de 3'5 días. **(a)** ¿Podemos afirmar que esta diferencia se debe al azar con una confianza del 88 %? **(b)** Con el mismo nivel de confianza, ¿cambiaría la respuesta si esta media de 3'5 días se hubiera obtenido al analizar el tiempo de estancia de 100 viajeros elegidos al azar? ***Solución: (a) Como*** $\mathbf{3'5} \in (\mathbf{3'45}, \mathbf{3'95})$ ***podemos afirmar que el tiempo de estancia medio es de 3'7 días y que la diferencia con respecto a la muestra es debida al azar. (b) Como*** $\mathbf{3'5} \notin (\mathbf{3'53}, \mathbf{3'87})$ ***rechazamos la hipótesis nula.***

13. Se ha llevado a cabo un estudio en diferentes países de la UE del porcentaje de la población que accede a la enseñanza superior. En los países escogidos se han obtenido los valores siguientes (medidos en tanto por ciento):

23'5 35'0 29'5 31'0 23'0 33'5 27'0 28'0 30'5

Se supone que estos porcentajes siguen una distribución normal con desviación típica igual al 5 %. Se desea contrastar con un nivel de significación del 5 % si los datos anteriores son compatibles con un valor medio del porcentaje de la población que cursa estudios superiores igual al 28 %. ***Solución: Como*** $\mathbf{29} \in (\mathbf{24'73}, \mathbf{31'26})$ ***podemos aceptar la hipótesis nula con una probabilidad de equivocarnos del 5%.***

14. Para una operación de compraventa de un supermercado se tiene, entre otras, la siguiente información: los vendedores afirman que la caja media por cliente es de 7'50 € por operación, con distribución normal. La empresa compradora efectuó un muestreo de tamaño 36 que dio un gasto medio de 7'22 € y una desviación típica de 0'56 €. **(a)**Para un nivel de significación del 5 %, indica si el muestreo es representativo en ensayo bilateral de la población que indican los vendedores. **(b)** En el supuesto de que los 7'50 € sea el valor mínimo del gasto de los clientes, comprueba la validez de la muestra en ensayo unilateral con el mismo nivel. ***Solución: (a) Como*** $\mathbf{7'22} \notin (\mathbf{7'317}, \mathbf{7'683})$ ***rechazamos la hipótesis nula. Por tanto el muestreo no es representativo. (b) Como*** $\mathbf{7'22} \notin (\mathbf{7'346}, +\infty)$ ***rechazamos la hipótesis nula. Por tanto el muestreo no es representativo.***

15. Se sabe que la edad de los aspirantes a un puesto de trabajo en un determinado organismo oficial es una variable normal con desviación típica igual a 5. Se observa una muestra de 125 personas que se presentan a una prueba para optar a un puesto de trabajo en el citado organismo, obteniéndose una edad media igual a 22'3 años. **(a)** ¿Se puede afirmar, con un nivel de significación del 5 %, que es igual a 21 años la edad media de los aspirantes al puesto de trabajo? **(b)** ¿Se puede afirmar, si el nivel de significación es del 1 %, que dicha edad media es menor o igual que 22? ***Solución: (a) Como*** $\mathbf{22'3} \notin (\mathbf{20'13}, \mathbf{21'87})$ ***rechazamos la hipótesis nula. Por tanto, no podemos afirmar, con un nivel de significación del 5 %, que la edad media de los aspirantes sea 21 años. (b) Como*** $\mathbf{22'3} \in (-\infty, \mathbf{23'04})$ ***aceptamos la hipótesis nula.***

16. Al lanzar 5000 veces una moneda al aire salieron 3000 caras. Se puede aceptar, con un nivel de significación del 5 % que la moneda no está trucada? ***Solución: Como*** $\mathbf{0'6} \notin (\mathbf{0'486}, \mathbf{0'514})$ ***rechazamos la hipótesis nula. Por tanto, la moneda está trucada.***

17. Una empresa dedicada a la fabricación de luminosos publicitarios anuncia que, como máximo, hay un 1 % de luminosos defectuosos. Se selecciona una muestra de 100 rótulos publicitarios y se observa que aparecen 3 defectuosos. **(a)** Con un nivel de significación del 5 %, ¿podemos aceptar la hipótesis? **(b)** ¿y con un nivel de confianza de 99 %? ***Solución: (a) Como*** $\mathbf{0'03} \notin (-\infty, \mathbf{0'026})$ ***rechazamos la hipótesis del fabricante. (b) Como*** $\mathbf{0'03} \in (-\infty, \mathbf{0'033})$ ***aceptamos la hipótesis nula, es decir, la del fabricante.***

Modelos resueltos

18. Una encuesta hecha a 400 familias de una gran ciudad dio un gasto medio anual en zapatos de 740 € por familia. Por estudios previos es conocida la desviación típica de la población $\sigma = 400$ €.

(a) Establece un intervalo de confianza del 95 % para estimar el gasto medio anual en zapatos por familia en esa ciudad. ***(b)*** Determina el error en la estimación del gasto medio si la confianza que establecemos es del 99 ***%. (c)*** ¿Con qué grado de confianza puede afirmarse que el gasto medio anual en zapatos por familia en esa ciudad está entre 710 € y 770 €?

Población (X="gasto en zapatos", μ desconocida, σ= 400 €)

Muestra (n = 400 > 30 $\Rightarrow \overline{X} = N(\mu_{\bar{x}} = \mu,\ \sigma_{\bar{x}} = \frac{400}{\sqrt{400}} = 20),\ \overline{x_i} = 740$€)

(a) Intervalo de confianza para μ: $\overline{x_i} \pm Z_{\frac{\alpha}{2}} \cdot \sigma_{\bar{x}}$

Como $1-\alpha = 0'95 \Rightarrow Z_{\frac{\alpha}{2}} = 1'96$

$$(740 - 1'96 \cdot 20, 740 + 1'96 \cdot 20) = (700'8,\ 779'2)$$

$\mu \in (700'8,\ 779'2)$ con una confianza del 95 %

(b) $\left.\begin{array}{c} Error = Z_{\frac{\alpha}{2}} \cdot \sigma\overline{x} \\ 1-\alpha = 0'99 \Rightarrow Z_{\frac{\alpha}{2}} = 2'575 \end{array}\right\} \Rightarrow E = 2'575 \cdot 20 = 51'5$ €. Cometemos un error máximo de 51'5 € en la estimación de la media de la población con una confianza del 99 %.

(c) Intervalo de confianza= $(710, 770) \Rightarrow Z_{\frac{\alpha}{2}} \cdot \sigma_{\bar{x}} = \frac{770-710}{2} = 30 \Rightarrow Z_{\frac{\alpha}{2}} \cdot 20 = 30 \Rightarrow Z_{\frac{\alpha}{2}} = 1'5$

$$P(Z < 1'5) = 1 - \alpha + \tfrac{\alpha}{2} \Rightarrow 0'9332 = 1 - \alpha + \tfrac{\alpha}{2} \Rightarrow \alpha = 0'1336 \Rightarrow 1 - \alpha = 0'8664$$

La confianza es del 86'64 %.

19. Es sabido que el 58 % de los votantes de un cierto distrito electoral apoyan al partido A.

(a)¿Cuál es la probabilidad de que una muestra aleatoria simple de 100 votantes de ese distrito de una proporción de simpatizantes de A del 60 % o menos?

(b) Supongamos ahora que se toma una muestra de 400 votantes del distrito. ¿Cuál es la probabilidad de que la proporción muestral sea del 60 % o menos?

(c) Sabiendo que si tomamos una muestra de 400 votantes la probabilidad de que la proporción de simpatizantes de A sea menor de L es 0’85, determina dicho valor L.

Población (p= 0’58 = proporción de votantes que apoyan a A)

(a) Muestra (n=100 >30 $\Rightarrow P = N(\mu_p = p = 0'58, \sigma_p = \sqrt{\frac{0'58 \cdot 0'42}{100}} = 0'0494))$

$P(P \le 0'6) = P\left(Z \le \frac{0'6 - 0'58}{0'0494}\right) = P(Z \le 0'41) = 0'6591$

(b) Muestra (n=400 >30 $\Rightarrow P = N(\mu_p = p = 0'58, \sigma_p = \sqrt{\frac{0'58 \cdot 0'42}{400}} = 0'0247))$

$P(P \le 0'6) = P\left(Z \le \frac{0'6 - 0'58}{0'0247}\right) = P(Z \le 0'81) = 0'7910$

(c) Muestra (n=400 >30 $\Rightarrow P = N(\mu_p = p = 0'58, \sigma_p = \sqrt{\frac{0'58 \cdot 0'42}{400}} = 0'0247))$

$P(P \le L) = 0'85 \Rightarrow P(Z \le \frac{L - 0'58}{0'0247}) = 0'85 \Rightarrow$ Interpolando obtenemos $\frac{L - 0'58}{0'0247} = 1'037 \Rightarrow L = 0'6056$

20. En un país se selecciona aleatoriamente una muestra de 900 personas. A la salida de los colegios electorales se les preguntó si habían votado al partido político X y 288 contestaron que sí y el resto que no.

(a) Determina un intervalo que nos dé el porcentaje de votos del partido X con un nivel de confianza del 90 %, explicando los pasos seguidos para su obtención.

(b) ¿De qué tamaño hemos de elegir la muestra si queremos que el error no exceda en 0’05 ?

Muestra ($p_i = \frac{288}{900} = 0'32,\ n = 900 > 30 \Rightarrow P = N\left(\mu_p = p, \sigma_p = \sqrt{\frac{0'32 \cdot 0'68}{900}} = 0'0155\right)$

Población: p a estimar.

(a) $1 - \alpha = 0'90 \Rightarrow Z_{\frac{\alpha}{2}} = 1'645$

Intervalo de confianza para p: $p_i \pm Z_{\frac{\alpha}{2}} \cdot \sigma_p = 0'32 \pm 1'645 \cdot 0'0155 \Rightarrow$

p se encuentra en el intervalo : $(0'2945, 0'3455)$ con una confianza del 90 %.

(b) $\left.\begin{array}{c} E = Z_{\frac{\alpha}{2}} \cdot \sigma_p = 0'05 \\ 1 - \alpha = 0'90 \Rightarrow Z_{\frac{\alpha}{2}} = 1'645 \end{array}\right\} \Rightarrow 0'05 = 1'645 \cdot \frac{\sqrt{0'32 \cdot 0'68}}{\sqrt{n}} \Rightarrow n = 235'5.. \simeq 236$

21. La presión arterial de los individuos de una población sigue un ley normal de media 120 y desviación típica 40. Resuelve las siguientes cuestiones:

(a) Calcula la probabilidad de que una persona elegida al azar tenga una tensión arterial superior a 140.

(b) Se considera una muestra de tamaño 36, ¿cuál es la probabilidad de que la media de la muestra tenga un valor superior a 140?

(c) Si conociéramos únicamente la desviación típica de la presión arterial $\sigma = 40$, ¿cuál ha de ser el tamaño de la muestra elegida para que el error en la estimación de la media poblacional sea menor o igual que 3 con una confianza del 95 % ?

Población: X="presión arterial" = $N(\mu = 120, \sigma = 40)$

(a) $P(X > 140) = P(Z > 0'5) = 1 - P(Z < 0'5) = 1 - 0'6915 = 0'3085$

(b) Muestra n=36>30 $\Rightarrow \overline{X} =$ "*presión media*" $= N(\mu_{\bar{x}} = 120, \sigma_{\bar{x}} = \frac{40}{\sqrt{36}} = \frac{20}{3})$

$P(\overline{X} > 140) = P(Z > 3) = 1 - P(Z < 3) = 1 - 0'9987 = 0'0013$

(c) Muestra n>30 $\Rightarrow \overline{X} =$ "*presión media*" $= N(\mu_{\bar{x}} = 120, \sigma_{\bar{x}} = \frac{40}{\sqrt{n}})$

$Error = 3 = Z_{\frac{\alpha}{2}} \cdot \sigma_{\bar{x}} = 1'96 \cdot \frac{40}{\sqrt{n}} \Rightarrow n = \left(\frac{1'96 \cdot 40}{3}\right)^2 = 682'9.. \simeq 683$

22. Una fábrica produce ciertas piezas de precisión con un porcentaje de defectuosas en torno al 10 %, aproximadamente. Seleccionamos una muestra de 36 piezas. Si la probabilidad de que la proporción de piezas defectuosas en la muestra sea menor que D es de 0'53, calcula el valor de D.

- Población: p=proporción de piezas defectuosas =0'1

Muestra (n=36 >30 $\Rightarrow P = N(\mu_p = p = 0'1, \sigma_p = \sqrt{\frac{0'1 \cdot 0'9}{36}} = 0'05))$

$P(P < D) = 0'53 \Rightarrow P(Z < \frac{D - 0'1}{0'05}) = 0'53 \Rightarrow Interpolando\ obtenemos\ \frac{D - 0'1}{0'05} = 0'075$

$\Rightarrow D = 0'05 \cdot 0'075 + 0'1 = 0'10375$

23. En una encuesta realizada sobre el examen de selectividad a 150 alumnos, estaban en contra de la prueba 105.

(a) ¿Entre qué valores podemos asegurar que está el porcentaje de alumnos contrarios a la selectividad con un nivel de confianza del 90 % ?

(b) ¿Entre qué valores podemos asegurar que está el porcentaje de alumnos contrarios a la selectividad con un nivel de confianza del 99 % ?

- Población (p= proporción de alumnos que están en contra de la prueba , a estimar)

Muestra ($p_i = \frac{105}{150} = 0'7,\ n = 150 > 30 \Rightarrow P = N\left(\mu_p = p, \sigma_p = \sqrt{\frac{0'7 \cdot 0'3}{150}} = 0'0374\right)$

(a) Intervalo de confianza para p : $p_i \pm Z_{\frac{\alpha}{2}} \cdot \sigma_p = 0'7 \pm Z_{\frac{\alpha}{2}} \cdot 0'0374$

Si $1 - \alpha = 0'90 \Rightarrow Z_{\frac{\alpha}{2}} = 1'645 \Rightarrow$ Intervalo de confianza para p:

$(0'7 - 1'645 \cdot 0'0374, 0'7 + 1'645 \cdot 0'0374) = (0'6385, 0'7615)$

(b) Si $1 - \alpha = 0'99 \Rightarrow Z_{\frac{\alpha}{2}} = 2'575 \Rightarrow$ Intervalo de confianza para p:

$(0'7 - 2'575 \cdot 0'0374, 0'7 + 2575 \cdot 0'0374) = (0'6106, 07894)$

Se puede observar que cuanto mayor es la confianza (manteniendo el tamaño de la muestra) el intervalo en el que se encuentra **p** tiene una amplitud mayor.