13 El muestreo estadístico

ACTIVIDADES INICIALES

13.I. Las notas obtenidas en matemáticas por 12 estudiantes de 2.º de Bachillerato son:

9, 5, 3, 9, 0, 10, 2, 1, 8, 9, 7, 6

a) Calcula la media.

b) Halla la varianza y la desviación típica.

c) Calcula la cuasivarianza.

a) $\overline{x} = \dfrac{0+1+2+3+5+6+7+8+9\cdot 3+10}{12} = 5{,}75$

b) $\sigma^2 = \dfrac{(0-5{,}75)^2+(1-5{,}75)^2+(2-5{,}75)^2+(3-5{,}75)^2+(5-5{,}75)^2+(6-5{,}75)^2+(7-5{,}75)^2+(8-5{,}75)^2+(9-5{,}75)^2\cdot 3+(10-5{,}75)^2}{12}$

$\sigma^2 = 11{,}19$

$\sigma = \sqrt{11{,}19} = 3{,}34$

c) $\hat{s}^2 = \dfrac{N}{N-1}s^2 = \dfrac{12}{12-1}11{,}19 = 12{,}20$

13.II. Lanza un dado 20 veces y anota los resultados de la cara superior.

a) Calcula la media y la varianza de los resultados obtenidos.

b) Halla la cuasivarianza.

La solución depende de lo que haya salido en los 20 lanzamientos a cada alumno.

13.III. Repite la actividad anterior, pero realizando 50 lanzamientos.

Ocurre lo mismo que en la actividad anterior.

EJERCICIOS PROPUESTOS

13.1. El consorcio de transporte de una comunidad quiere saber la opinión que tienen los ciudadanos sobre el servicio de cercanías. Para ello, unos encuestadores realizan una serie de entrevistas a los viajeros que acceden a este servicio en tres estaciones de la red.

a) ¿Cuál es la población?

b) ¿Cuál es la muestra?

a) Todos los ciudadanos que utilizan el servicio de cercanías.

b) Los viajeros de las tres estaciones de la red que son encuestados.

13.2. En los años treinta, en Estados Unidos se hizo una encuesta telefónica para pronosticar el ganador de las siguientes elecciones presidenciales. El pronóstico fue que ganaría el candidato republicano, pero en realidad ganó el candidato demócrata.

¿Crees que la muestra elegida fue representativa? ¿Por qué?

La muestra seleccionada no fue representativa porque no todos los individuos de la población tenían la misma probabilidad de ser elegidos: algunos no tenían teléfono, en algunas ciudades había más teléfonos que en otras…

Esto conlleva una muestra que no estaba formada por los distintos individuos que componían la población de Estados Unidos, de modo que no es raro que el pronóstico no fuera acertado.

13.3. El 5% de los pasteles que hace un pastelero tienen un exceso de peso. Se toma una muestra de 45 pasteles:

a) ¿Cuál es la distribución que sigue la proporción de pasteles con exceso de peso de la muestra?

b) Halla la probabilidad de que en la muestra existan al menos cuatro pasteles con exceso de peso.

a) $\sigma = \sqrt{\dfrac{0,05(1-0,05)}{45}} = 0,0325 \Rightarrow N(0,05;\ 0,0325)$

b) $\dfrac{4}{45}100 = 8,89\% \Rightarrow P(\hat{P} \geq 0,09) = P\left(Z \geq \dfrac{0,09-0,05}{0,0325}\right) = P(Z \geq 1,23) = 1 - P(Z < 1,23) = 1 - 0,8907 = 0,1093$

13.4. En la elección para formar parte del consejo escolar, un alumno ha recibido un 50% de votos desfavorables, si se elige una muestra de 40 alumnos que han votado.

a) ¿Cuál es la distribución que sigue la proporción de votantes que le han votado?

b) Halla la probabilidad de que más del 40% de la muestra le votasen.

a) $\sigma = \sqrt{\dfrac{0,5(1-0,5)}{40}} = 0,079 \Rightarrow N(0,5;\ 0,079)$

b) $P(\hat{P} \geq 0,4) = P\left(Z \geq \dfrac{0,4-0,5}{0,079}\right) = P(Z \geq -1,27) = P(Z < 1,27) = 0,8980$

13.5. La emisión de óxido de nitrógeno de los vehículos de cierta marca sigue una distribución normal con media μ = 1,2 y desviación típica σ = 0,4. Se escoge al azar una muestra de 25 vehículos.

a) ¿Cuál es la distribución en el muestreo de la media?

b) Halla la probabilidad de que la media de la muestra sea mayor de 1,2.

a) $\sigma = \dfrac{0,4}{\sqrt{25}} = 0,08 \Rightarrow N(1,2;\ 0,08)$

b) $P(\overline{X} > 1,2) = 0,5$, ya que $x = 1,2$ (la media) es el eje de simetría de la gráfica de la función normal.

13.6. El peso de las vacas de una determinada ganadería se distribuye según una normal de media 495 kg y desviación típica 44 kg. Se toma una muestra de 35 vacas de esa ganadería. Halla la probabilidad de que la media muestral:

a) Sea mayor de 500 kg.

b) Sea menor de 480 kg.

c) Esté comprendida entre 490 y 500 kg.

a) $\sigma = \dfrac{44}{\sqrt{35}} = 7,44 \Rightarrow N(495;\ 7,44) \Rightarrow P(X > 500) = P\left(Z > \dfrac{500-495}{7,94}\right) = P(Z > 0,67) = 1 - P(Z \leq 0,67) = 0,2514$

b) $P(\overline{X} < 480) = P\left(Z < \dfrac{480-495}{7,94}\right) = P(Z < -2,02) = P(Z \geq 2,02) = 1 - P(Z < 2,02) = 1 - 0,9783 = 0,0217$

c) $P(490 < \overline{X} < 500) = P\left(\dfrac{490-495}{7,94} < Z < \dfrac{500-495}{7,94}\right) = P(-0,63 < Z < 0,63) = P(Z < 0,63) - P(Z < -0,63) =$

$= 2P(Z < 0,63) - 1 = 2 \cdot 0,6357 - 1 = 0,2714$

13.7. Las notas de PAU de los estudiantes de una localidad tienen una media de μ = 5,35 y desviación típica σ = 1,26. Se toma al azar una muestra de 100 estudiantes.

¿Cuál es la distribución que sigue la suma de las notas de la muestra?

$N(n\mu,\ \sigma\sqrt{n}) = N(100 \cdot 5,35;\ 1,26 \cdot 10) = N(535;\ 12,6)$

13.8. Las consultas de un médico de cabecera duran una media de 8 minutos, con una desviación típica de 2,3 minutos. Si una tarde tiene citados 32 pacientes, ¿cuál es la probabilidad de que los atienda en menos de 4 horas?

$N(32\cdot 8;\ 2{,}3\sqrt{32}) = N(256;\ 13{,}0107)$

$P(T < 240) = P\left(Z < \frac{240-256}{13{,}01}\right) = P(Z < -1{,}23) = P(Z > 1{,}23) = 1 - P(Z \le 1{,}23) = 1 - 0{,}8907 = 0{,}1093$

13.9. Una empresa comercializa sal y la empaqueta en bolsas de 500 gramos. Se sabe que los pesos reales de las bolsas siguen una distribución normal de media 498 g y desviación típica 8 g.

a) Si se toma una muestra de 100 bolsas, ¿qué probabilidad hay de que el peso total de las mismas sea inferior a 48 kg?

b) Si se toma una muestra de 50 bolsas, ¿cuál es la probabilidad de que el peso total de las mismas supere los 25 kg?

a) $N(498 \cdot 100,\ 8 \cdot \sqrt{100}) = N(49\,800,\ 80)$

$P(T < 48000) = P\left(Z < \frac{48000-49800}{80}\right) = P(Z < -2{,}25) = P(Z \ge 22{,}5) = 0$

b) $N(498 \cdot 50,\ 8 \cdot \sqrt{50}) = N(24\,900;\ 56{,}57)$

$P(T > 25000) = P\left(Z > \frac{25000-24900}{56{,}57}\right) = P(Z > 1{,}76) = P(Z \le 1{,}76) = 1 - 0{,}9608 = 0{,}0392$

13.10. Se sabe que el peso *X* de la grasa corporal en adultos que no hacen ejercicio sigue una distribución con media de 24,3 kg y desviación típica de 2,4. En cambio, el peso *Y* de la grasa en adultos que hacen ejercicio regularmente se distribuye con una media de 20,1 y desviación típica de 1,7. Si se eligen en ambas poblaciones muestras aleatorias de 50 elementos, ¿cuál es la probabilidad de que la diferencia de la grasa corporal media sea mayor de 3 kg?

$N\left(24{,}3 - 10{,}1;\ \sqrt{\frac{2{,}4^2}{50} + \frac{1{,}7^2}{50}}\right) = N(4{,}2;\ 0{,}4159) \Rightarrow$

$\Rightarrow P(\overline{X}_1 - \overline{X}_2 > 3) = P\left(Z > \frac{3-4{,}2}{0{,}4159}\right) = P(Z > -2{,}89) = P(Z < 2{,}89) = 0{,}9981$

13.11. Uno de los principales fabricantes de televisores compra piezas a dos compañías. Las piezas de la compañía *A* tienen una vida media de 7,2 años con una desviación estándar de 0,8 años, mientras que las de la *B* tienen una vida media de 6,7 años con una desviación estándar de 0,7. Determina la probabilidad de que una muestra aleatoria de 34 piezas de la compañía *A* tenga una vida media de al menos un año más que la de una muestra aleatoria de 40 piezas de la compañía *B*.

$N\left(7{,}2 - 6{,}7;\ \sqrt{\frac{0{,}8^2}{34} + \frac{0{,}7^2}{40}}\right) = N(0{,}5;\ 0{,}1763)$

$P(\overline{X}_1 - \overline{X}_2 > 1) = P\left(Z > \frac{1-0{,}5}{0{,}1763}\right) = P(Z > 2{,}84) = 1 - P(Z < 2{,}84) = 1 - 0{,}9977 = 0{,}0023$

13.12. Una variable aleatoria tiene media μ = 30 y desviación típica σ = 3,5.

Se eligen al azar muestras de tamaño *n*.

¿Qué se puede decir de la distribución de las medias muestrales en los siguientes casos?

a) *n* = 20

b) *n* = 40

a) No se podría decir nada porque $n < 30$ y no se sabe si la población de partida es normal; por tanto, no se puede aplicar el teorema central del límite.

b) $N\left(30;\ \frac{3,5}{\sqrt{40}}\right) = N(30;\ 0,5534)$

13.13. En la clase de 2.º de Bachillerato se sabe que el peso de los alumnos se distribuye normalmente con media de 64 kg y desviación típica de 8 kg. Se toma una muestra de 10 alumnos.

a) ¿Cuál es la distribución que siguen las medias muestrales?

b) ¿Cuál es la probabilidad de que la media del peso en la muestra esté entre 60 y 68 kg?

a) $N\left(64;\ \frac{8}{\sqrt{10}}\right) = N(64;\ 2,5298)$

b) $P(60 \le \overline{X} \le 68) = P\left(\frac{60-64}{2,5298} \le Z \le \frac{68-64}{2,5298}\right) = P(-1,58 \le Z \le 1,58) = P(Z \le 1,58) - P(Z \le -1,58) =$

$= 2P(Z \le 1,58) - 1 = 0,8858$

EJERCICIOS

Población y muestra

13.14. Antes de unas elecciones, un periódico llama telefónicamente y de forma aleatoria a 2000 ciudadanos con derecho a voto y les pregunta por el candidato al que van a votar. Así, espera obtener una estimación fiable de los resultados que se van a producir:

a) ¿Cuál es la población?

b) ¿Cuál es la muestra?

c) ¿Es realmente aleatoria la muestra?

d) ¿Se puede calcular la media muestral? ¿Por qué?

a) Los ciudadanos con derecho al voto.

b) Los 2000 ciudadanos elegidos.

c) No, ya que no todo el mundo tiene teléfono y, además, hay ciudadanos con más de un teléfono.

d) No, ya que es una variable cualitativa.

Tipos de muestreo

13.15. La tecla $\boxed{\text{RAN}}$ de las calculadoras científicas nos permite obtener números aleatorios comprendidos entre 0,000 y 0,999. Así, si tenemos una población de *N* elementos y multiplicamos un número aleatorio obtenido con dicha tecla por *N*, obtendremos números comprendidos entre 0 y *N*. Por tanto, si pulsamos la secuencia de teclas $\boxed{\text{RAN}}\ \boxed{\text{x}}\ \boxed{\text{N}}\ \boxed{+}\ \boxed{1}$ y nos quedamos con la parte entera del número, iremos obteniendo aleatoriamente los elementos que integrarán la muestra. Utiliza este método para obtener una muestra aleatoria de tamaño 10 de una población de 450 individuos.

Respuesta libre.

13.16. Alberto trabaja en un tren revisando que los viajeros llevan el billete correcto. Como hoy el tren va totalmente lleno, 300 viajeros, no puede comprobar que todos los viajeros llevan el billete correcto. Por ello va a revisar el billete a 75 viajeros que los elegirá mediante un muestreo sistemático. Explica cómo lo hará.

Numerará a los viajeros del 1 al 300. Como $\frac{300}{75} = 4$, elegirá un viajero al azar e irá revisando de cuatro en cuatro. Por ejemplo, si empieza por el 155, revisará a los viajeros 155, 159, 163, 167, etc.
Obviamente, después del 300 va el 1, 2, 3, etc. Así revisaría a los 75 viajeros que pretendía.

13.17. En un reportaje de investigación se quieren conocer las opiniones sobre unos temas de personas que sean drogodependientes. ¿Qué técnica de muestreo recomendarías para seleccionar la muestra?

Como es un colectivo difícil de localizar, la mejor técnica de muestreo es la bola de nieve.

Distribuciones en el muestreo de una proporción

13.18. En unas elecciones generales, el presidente del gobierno elegido por los ciudadanos ha recibido un 45% de los votos favorables. Se escoge una muestra al azar de 50 votantes.
a) ¿Cuál es la distribución que sigue la proporción de votantes que han votado al presidente de la muestra?
b) Halla la probabilidad de que más de la mitad de los votantes de la muestra votasen al presidente.

a) $N\left(p; \sqrt{\frac{p(1-p)}{n}}\right) = N\left(0,45; \sqrt{\frac{0,45 \cdot 0,55}{50}}\right) = N(0,45; 0,070)$

b) $P(\hat{P} > 0,5) = P\left(Z > \frac{0,5-0,45}{0,070}\right) = P(Z > -0,71) = P(Z < 0,71) = 0,2388$

13.19. El 4% de las piezas que produce una máquina son defectuosas. Se toma una muestra aleatoria de 80 piezas:
a) ¿Cuál es la distribución que sigue la proporción de piezas defectuosas de la muestra?
b) Halla la probabilidad de que en la muestra existan menos de 3 piezas defectuosas.

a) $N\left(p; \sqrt{\frac{p\,(1-p)}{n}}\right) = N\left(0,04; \sqrt{\frac{0,04 \cdot 0,96}{80}}\right) = N(0,04; 0,022)$

b) Tres piezas de 80 da una proporción de $\frac{3}{80} = 0,0375$. La probabilidad pedida es:

$$P(\hat{P} < 0,0375) = P\left(Z < \frac{0,0375 - 0,04}{0,022}\right) = P(Z < -0,11) = P(Z > 0,11) = 1 - P(Z < 0,11) = 0,4562$$

Distribuciones en el muestreo de la media

13.20. Una población de un tipo de plantas tiene una talla media de 15 cm y desviación típica de 2,5 cm. Se toma al azar una muestra de 45 plantas. ¿Cuál es la probabilidad de que la media de las tallas de la muestra sea superior a 12,5 cm?

La media de las tallas sigue una distribución $N\left(\mu; \frac{\sigma}{\sqrt{n}}\right) = N\left(15; \frac{2,5}{\sqrt{45}}\right) = N(15; 0,37)$.

La probabilidad pedida es:

$$P(\overline{X} > 12,5) = P\left(Z > \frac{12,5 - 15}{0,37}\right) = P(Z > -6,76) = P(Z < 6,76) = 1$$

Distribuciones en el muestreo de la suma de muestras

13.21. Se sabe que el peso medio de los pasajeros de un avión es 74 kg, con una desviación típica de 6 kg. Por las normativas de seguridad, la suma de los pesos de los pasajeros no puede superar las 3 toneladas. Si la compañía aérea ha vendido 40 pasajes, ¿cuál es la probabilidad de que cumpla con esa normativa de seguridad?

La suma de los pesos sigue una distribución $N(n\mu, \sigma\sqrt{n}) = N(40 \cdot 74, 6\sqrt{40}) = N(2960; 37{,}95)$.

La probabilidad pedida es:

$$P(T \leq 3000) = P\left(Z \leq \frac{3\,000 - 2\,960}{37{,}95}\right) = P(Z \leq 1{,}05) = 0{,}8531$$

13.22. El voltaje de las baterías de un fabricante es de 15 voltios y desviación típica 0,2 voltios. Si se eligen al azar 40 de esas baterías y se conectan en serie, lo que supone que se suman los voltajes, calcula la probabilidad de que tengan un voltaje conjunto de más de 60,8 voltios.

El voltaje conjunto de las 40 baterías sigue una distribución $N(n\mu, \sigma\sqrt{n}) = N(40 \cdot 15; 0{,}2\sqrt{40}) = N(60; 0{,}4)$.

La probabilidad pedida es:

$$P(T > 60{,}8) = P\left(Z > \frac{60{,}8 - 60}{0{,}4}\right) = P(Z > 2) = 1 - P(Z \leq 2) = 1 - 0{,}9772 = 0{,}0228$$

Distribuciones en el muestreo de la diferencia de las medias

13.23. Una empresa de telefonía móvil dota a sus teléfonos de baterías que provienen de dos fábricas *A* y *B*. Las baterías que provienen de la fábrica *A* duran una media de 6,1 días, con una desviación típica de 0,4 días, mientras que las baterías que provienen de la fábrica *B* duran una media de 5,4 días, con una desviación típica de 0,6 días. Se toma una muestra al azar formada por 42 baterías de la fábrica *A* y 36 de la fábrica *B*.
a) ¿Cuál es la distribución de la diferencia de medias para la muestra anterior?
b) ¿Cuál es la probabilidad de que la diferencia de las medias sea mayor de 0,5 días?

Baterías $A \rightarrow N(6{,}1; 0{,}4)$
Baterías $B \rightarrow N(5{,}4; 0{,}6)$

a) La distribución que sigue es $N\left(\mu_1 - \mu_2; \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}\right) = N\left(6{,}1 - 5{,}4; \sqrt{\frac{0{,}4^2}{42} + \frac{0{,}6^2}{36}}\right) = N(0{,}7; 0{,}12)$.

b) $P(\overline{X}_1 - \overline{X}_2 > 0{,}5) = P\left(Z > \frac{0{,}5 - 0{,}7}{0{,}12}\right) = P(Z > -1{,}67) = P(Z < 1{,}67) = 0{,}9525$

Teorema central del límite

13.24. El dinero que se gastan los adolescentes de entre 16 y 18 años durante un fin de semana sigue una distribución desconocida de media 6,20 € y desviación típica 1,90 €. Se toma al azar una muestra de 60 de esos adolescentes.
a) ¿Qué distribución sigue la media del gasto en dicha muestra?
b) ¿Cuál es la probabilidad de que la media del gasto en esa muestra sea superior a 7 €?
c) ¿Qué hubiese ocurrido si tomamos una muestra de solo 28 estudiantes?

a) Sigue aproximadamente una normal $N\left(6{,}2; \frac{1{,}9}{\sqrt{60}}\right) = N(6{,}2; 0{,}25)$.

b) $P(\overline{X} > 7) = P\left(Z > \frac{7 - 6{,}2}{0{,}25}\right) = P(Z > 3{,}2) \approx 0$

c) Con 28 alumnos, como la distribución de partida no es normal, no podemos asegurar que $\overline{X}$ sigue una normal.

PROBLEMAS

13.25. En un instituto hay 600 alumnos, 70 profesores y 10 auxiliares. Si queremos seleccionar una muestra de 68 individuos con la técnica del muestreo estratificado, ¿cómo deberíamos proceder?

En total hay 600 + 70 + 10 = 680 posibles integrantes de la muestra. Escogeremos al azar n_1 alumnos, n_2 profesores y n_3 auxiliares, donde n_1, n_2 y n_3 se calculan así:

$\frac{68}{680} = \frac{n_1}{600} \Rightarrow n_1 = 60$ alumnos $\qquad \frac{68}{680} = \frac{n_2}{70} \Rightarrow n_2 = 7$ profesores

$\frac{68}{680} = \frac{n_3}{10} \Rightarrow n_3 = 1$ auxiliar

Luego elegiremos al azar a 60 alumnos, 7 profesores y 1 auxiliar.

13.26. Consideremos una urna compuesta por 9 bolas numeradas del 1 al 9. Se extraen al azar y con reemplazamiento 40 bolas.

a) Halla la distribución de las sumas de las muestras de tamaño 40.

b) Calcula la probabilidad de que la suma de las puntuaciones sea mayor que 195.

a) La media de la población es $\mu = \frac{1+2+3+4+5+6+7+8+9}{9} = 5$, y la desviación típica,

$$\sigma^2 = \sqrt{\frac{(1-5)^2+(2-5)^2+(3-5)^2+(4-5)^2+(5-5)^2+(6-5)^2+(7-5)^2+(8-5)^2+9-5)^2}{9}} =$$

$$= \sqrt{\frac{16+9+4+1+0+1+4+9+16}{9}} = \sqrt{6{,}67} = 2{,}58.$$

La distribución de las sumas es: $N(n\mu, \sigma\sqrt{n}) = N(40 \cdot 5; 2{,}58\sqrt{40}) = N(200; 16{,}32)$.

b) $P(T > 195) = P\left(Z > \frac{195-200}{16{,}32}\right) = P(Z > -0{,}31) = P(Z < 0{,}31) = 0{,}6217$

13.27. La talla de los soldados de un ejército sigue una distribución normal *N*(175; 8,5). Para realizar una determinada tarea se elige al azar a 68 soldados. Calcula la probabilidad de que la talla media de esos soldados sea superior a 173 cm.

$$N\left(175; \frac{8{,}5}{\sqrt{68}}\right) = N(175; 1{,}03) \Rightarrow P(\overline{X} > 173) = P\left(Z > \frac{173-175}{1{,}03}\right) = P(Z > -1{,}94) = P(Z < 1{,}94) = 0{,}9738$$

13.28. La duración media de un tipo *A* de pilas eléctricas alcalinas es de μ_A = 65 h, con una desviación típica de σ_A = 3 h. Mientras, la duración media de otro tipo de pilas no alcalinas *B* es de μ_B = 50 h, con una desviación típica de σ_B = 2,5 h. Se toma al azar una muestra formada por 35 pilas tipo *A* y 80 pilas tipo *B*.

a) Halla la distribución en la muestra de la diferencia de medias.

b) Calcula la probabilidad de que la diferencia de medias en la muestra sea superior a 20 horas.

c) Calcula la probabilidad de que la diferencia de medias en la muestra sea inferior a 15 horas.

d) Calcula la probabilidad de que la diferencia de medias en la muestra esté comprendida entre 15 y 22 horas.

a) $N\left(\mu_A - \mu_B; \sqrt{\frac{\sigma_A^2}{n_A}+\frac{\sigma_B^2}{n_B}}\right) = N\left(65-50; \sqrt{\frac{3^2}{35}+\frac{2{,}5^2}{80}}\right) = N(15; 0{,}58)$

b) $P(\overline{X}_A - \overline{X}_B > 20) = P\left(Z > \frac{20-15}{0{,}58}\right) = P(Z > 8{,}62) \approx 0$

c) $P(\overline{X}_A - \overline{X}_B < 15) = P\left(Z < \frac{15-15}{0{,}58}\right) = P(Z < 0) = 0{,}5$

d) $P(15 \le \overline{X}_A - \overline{X}_B \le 22) = P\left(\frac{15-15}{0{,}58} \le Z \le \frac{22-15}{0{,}58}\right) = P(0 \le Z \le 12{,}07) = P(Z \le 12{,}07) - P(Z < 0) =$

$= 1 - 0{,}5 = 0{,}5$

13.29. En una gran empresa quieren averiguar si hay una gran diferencia entre los salarios de los hombres y de las mujeres. Para ello tomamos una muestra aleatoria de 50 hombres, y obtenemos que la media de sus salarios es μ_H = 1248 €, con una desviación típica de σ_H = 150 €, y tomamos una muestra aleatoria de 40 mujeres, y obtenemos que la media de sus salarios es μ_M = 1018 €, con una desviación típica de σ_M = 80 €.

a) Halla la distribución en el muestreo de la diferencia de medias.

b) ¿Cuál es la probabilidad de que la diferencia de las medias entre los salarios de los hombres y las mujeres sea superior a 200 €?

Hombres: $n_H = 50$; $\mu_H = 1248$ €; $\sigma_H = 150$ €
Mujeres: $n_M = 40$; $\mu_M = 1018$ €; $\sigma_M = 80$ €

a) $N\left(\mu_H - \mu_M;\ \sqrt{\dfrac{\sigma_H^2}{n_H} + \dfrac{\sigma_M^2}{n_M}}\right) = N\left(1\,248 - 1\,018;\ \sqrt{\dfrac{150^2}{50} + \dfrac{80^2}{40}}\right) = N(230; 24{,}70)$

b) $P(\overline{X}_H - \overline{X}_M > 200) = P\left(Z > \dfrac{200 - 230}{24{,}70}\right) = P(Z > -1{,}21) = P(Z < 1{,}21) = 0{,}8869$

PROFUNDIZACIÓN

13.30. Una prueba consta de 100 cuestiones. La puntuación media que se ha obtenido es de 68 puntos, con una desviación típica de 10 puntos.

a) ¿Cuál es la probabilidad de que de dos grupos formados por 40 y 36 personas, elegidas al azar de entre los que han realizado la prueba, difiera su puntuación media en más de 5 puntos?

b) ¿Y de que difiera entre 3 y 7 puntos?

La variable $\overline{X}_{40} - \overline{X}_{36}$ sigue una distribución $N\left(\mu_{40} - \mu_{36};\ \sqrt{\dfrac{\sigma_{40}^2}{n_{40}} + \dfrac{\sigma_{36}^2}{n_{36}}}\right) = N\left(0;\ \sqrt{\dfrac{10^2}{40} + \dfrac{10^2}{36}}\right) = N(0; 2{,}30)$.

a) $P(\overline{X}_{40} - \overline{X}_{36} > 5) = P\left(Z > \dfrac{5 - 0}{2{,}30}\right) = P(Z > 2{,}17) = 1 - P(Z < 2{,}17) = 1 - 0{,}9850 = 0{,}0150$

Observación: Si se considera que la probabilidad pedida es $P(|\overline{X}_{40} - \overline{X}_{36}| > 5)$, como la media es 0, sería 0,03.

b) $P(3 \le \overline{X}_{40} - \overline{X}_{36} \le 7) = P\left(\dfrac{3 - 0}{2{,}3} \le Z \le \dfrac{7 - 0}{2{,}3}\right) = P(1{,}30 \le Z \le 3{,}04) = P(Z \le 3{,}04) - P(Z \le 1{,}30) =$
$= 0{,}9988 - 0{,}9032 = 0{,}0956$

13.31. Un laboratorio fabrica comprimidos efervescentes en forma de disco, cuyo diámetro *X* quiere controlar. La variable aleatoria *X* tiene por media μ = 15 mm y desviación típica σ = 4 mm. A intervalos fijos de tiempo se extraen muestras de tamaño 64 de las que se miden los diámetros.

a) ¿Qué tipo de muestreo se está realizando?

b) Determina la distribución en el muestreo de la variable media muestral.

c) Calcula la probabilidad de que dicha variable esté comprendida entre 13 y 16 mm.

a) Muestreo aleatorio sistemático.

b) La media muestral $\overline{X}$ sigue en el muestreo una normal $N\left(15;\ \dfrac{4}{\sqrt{64}}\right) = N(15; 0{,}5)$.

c) $P(13 \le \overline{X} \le 16) = P\left(\dfrac{13 - 15}{0{,}5} \le Z \le \dfrac{16 - 15}{0{,}5}\right) = P(-4 \le Z \le 2) = P(Z \le 2) - P(Z \le -4) =$
$= P(Z \le 2) - P(Z \ge 4) = P(Z \le 2) - 1 + P(Z \le 4) = 0{,}9772 - 1 + 1 = 0{,}9772$

13.32. Las tallas de los jugadores de baloncesto de 16 años en España siguen una distribución normal *N*(179, 18). Se toma al azar una muestra de 100 jugadores de este tipo. Halla los límites inferior y superior del intervalo (179 – *k*, 179 + *k*) para que la media de muestra elegida esté en ese intervalo con una seguridad del 90%.

La distribución sigue una $N\left(179;\ \dfrac{18}{\sqrt{100}}\right) = N(179;\ 1{,}8)$.

$$P(179 - k \le \overline{X} \le 179 + k) = P\left(\frac{179 - k - 179}{1{,}8} \le Z \le \frac{179 + k - 179}{1{,}8}\right) = P\left(\frac{-k}{1{,}8} \le Z \le \frac{k}{1{,}8}\right) =$$

$$= P\left(Z \le \frac{k}{18}\right) - P\left(Z \le -\frac{k}{18}\right) = P\left(Z \le \frac{k}{18}\right) - P\left(Z \ge \frac{k}{18}\right) = P\left(Z \le \frac{k}{18}\right) - 1 + P\left(Z \le \frac{k}{18}\right) = 0{,}9 \Rightarrow$$

$$2P\left(Z \le \frac{k}{18}\right) = 1{,}9 \Rightarrow \Rightarrow P\left(Z \le \frac{k}{1{,}8}\right) = 0{,}95 \Rightarrow \frac{k}{1{,}8} = 1{,}645 \Rightarrow k = 1{,}645 \cdot 1{,}8 = 2{,}961$$

El intervalo buscado es (190 – 2,961; 190 + 2,961) = (187,039; 192,961).

13.33. En un aeropuerto operan dos compañías aéreas, *A* y *B*. El 60% de los vuelos los lleva a cabo la compañía *A*, y el 40% restante, la compañía *B*. La compañía aérea *A* sabe que los retrasos de sus vuelos se distribuyen según una normal de media 18 minutos y desviación típica 4 minutos. Mientras, en la compañía aérea *B*, los retrasos se distribuyen según una normal de media 10 minutos y desviación típica 3 minutos. Se toma una muestra de tamaño 100 al azar, con la técnica del muestreo estratificado proporcional. Halla la probabilidad de que la diferencia media de los retrasos de la muestra sea superior a 10 minutos.

Como la muestra se ha elegido mediante un muestreo estratificado proporcional, se han elegido 60 vuelos de la compañía *A* y 40 de la compañía *B*.

La distribución en el muestreo de la diferencia de medias sigue una normal $N\left(18 - 10;\ \sqrt{\dfrac{4^2}{60} + \dfrac{3^2}{30}}\right) = N(8;\ 0{,}75)$.

$$P(\overline{X}_A - \overline{X}_B > 10) = P\left(Z > \frac{10 - 8}{0{,}75}\right) = P(Z > 2{,}67) = 1 - P(Z < 2{,}67) = 1 - 0{,}9962 = 0{,}0038$$

RELACIONA Y CONTESTA

Elige la única respuesta correcta en cada caso:

13.1. Un grupo de música tiene compuestos 50 temas. Para tocar en directo debe elegir 20 de esas canciones.

Para ello, las numera del 1 al 50 y hace un sorteo mediante el cual obtiene las 20 canciones que tocará. La técnica que ha elegido el grupo es:

A) Muestreo aleatorio sistemático.

B) Muestreo aleatorio estratificado.

C) Muestreo aleatorio simple.

D) Muestreo por conglomerados y áreas.

E) Muestreo "bola de nieve".

La respuesta correcta es la C.

13.2. Un jugador de baloncesto tiene un acierto de un 40% en sus tiros de 3 puntos. Si en una temporada ha lanzado 95 tiros de 3 puntos, el muestreo de la proporción sigue aproximadamente una distribución:

A) *N*(38; 0,05)

B) *N*(0,4; 0,05)

C) *N*(38; 0,0025)

D) *N*(0,4; 0,0025)

E) No se puede aproximar por una normal.

Sigue una distribución $N\left(p;\ \sqrt{\dfrac{p(1-p)}{n}}\right) = N\left(0{,}4;\ \sqrt{\dfrac{0{,}4 \cdot 06}{95}}\right) = N(0{,}4\ ;\ 0{,}05)$. La respuesta correcta es la B.

13.3. Los pesos de los alumnos de un centro educativo siguen una distribución normal de parámetros μ = 164 cm y σ = 11 cm. Si se toma una muestra de 50 personas, la probabilidad de que la media de la muestra sea menor que 164,2 cm es aproximadamente:

A) 0,6026

B) 0,7454

C) 0,7764

D) 0,8508

E) 0,8166

La distribución en el muestreo de la media sigue una normal $\overline{X} = N\left(\mu ; \frac{\sigma}{\sqrt{n}}\right) = N(164 ; 0{,}22)$.

Por tanto, $P\left(\overline{X} < 164{,}2\right) = P\left(Z < \frac{164{,}2 - 164}{0{,}22}\right) = P\left(Z < 0{,}91\right) = 0{,}8186$

La respuesta correcta es la E.

13.4. Se les pasa una misma prueba de Inglés a los estudiantes de 2.º de Bachillerato de dos poblaciones *A* y *B*. En la población *A*, las puntuaciones siguen una distribución normal *N*(6,3; 1,5), mientras que las puntuaciones de la población *B* siguen una normal *N*(7; 2,1). Si se elige una muestra aleatoria de 40 estudiantes de la población *A* y 50 de la población *B*, la probabilidad de que la diferencia de las medias sea superior a 1 punto es aproximadamente:

A) 0,6517

B) 0,7852

C) 0,8438

D) 0,9345

E) 0,9772

El muestreo de las diferencias de las medias sigue una distribución normal

$N\left(\mu_1 - \mu_2 ; \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}\right) = N\left(0{,}7;\ 0{,}38\right)$.

Por tanto, $P\left(\overline{X}_A - \overline{X}_B > 1\right) = P\left(Z > \frac{1 - 0{,}7}{0{,}38}\right) = P\left(Z > 0{,}79\right) = 0{,}7852$

La respuesta correcta es la B.

Señala, en cada caso, las respuestas correctas:

13.5. Señala cuáles de los siguientes tipos de muestreos son aleatorios:

A) “Bola de nieve”

B) Muestreo estratificado

C) Muestreo por conglomerados y áreas

D) Muestreo por cuotas

E) Muestreo sistemático

B, C y E son los muestreos aleatorios que aparecen.

13.6. A medida que el tamaño de la muestra tiende a infinito:

A) La desviación típica de la distribución en el muestreo de la proporción tiende a 0.

B) La media de la distribución en el muestreo de la media tiende a infinito.

C) La desviación típica de la distribución en el muestreo de la media tiende a 0.

D) La media de la distribución en el muestreo de las sumas muestrales tiende a 0.

E) La desviación típica de la distribución en el muestreo de las sumas muestrales tiende a 0.

A y C son correctas.

B no es correcta, ya que la media coincide con la media poblacional.

D no es correcta porque tiende a infinito, siempre que la media poblacional no sea nula.

E no es correcta porque tiende a infinito.

Elige la relación correcta entre las dos afirmaciones dadas:

13.7. Sea *X* una variable aleatoria.

a) *X* no sigue una distribución normal.

b) Se toma una muestra de tamaño 25.

c) La distribución en el muestreo de las medias muestrales se puede aproximar por una normal.

A) a y b $\Rightarrow$ c

B) No a y b $\Rightarrow$ c

C) c $\Rightarrow$ No a

D) No b $\Rightarrow$ c

E) a y b $\Rightarrow$ No c

B y E son correctas.

A no es correcta, ya que la muestra es pequeña.

C no es correcta, ya que puede ser que la distribución de partida no sea normal y la muestra sea grande.

D no es cierta, ya que "No b" no quiere decir que la muestra sea grande, sino que la muestra no es de 25 elementos, es decir, que puede seguir siendo pequeña.

Señala los datos innecesarios para contestar:

13.8. Sea *X* una variable aleatoria que sigue una distribución normal de media μ y desviación típica σ. Se toma una muestra de tamaño *n*. Para saber si la distribución de las medias muestrales se puede aproximar por una normal:

A) Son imprescindibles todos los datos que se mencionan.

B) Solo se puede eliminar el dato σ que indica la desviación típica de la población de partida.

C) Se puede eliminar el dato *n* que indica el tamaño de la muestra.

D) Se puede eliminar el dato de que la población de partida es normal.

E) El único dato relevante es conocer si el tamaño de la muestra es grande ($n \geq 30$).

La respuesta correcta es la C.

Siendo la población de partida normal, la distribución en el muestreo de las medias muestrales será normal, independientemente del tamaño de la muestra y la media y la desviación típica de la población de partida.

Analiza si la información suministrada es suficiente para contestar la cuestión:

13.9. En un instituto se sabe que los pesos de las mochilas que llevan los alumnos tienen por media 6,3 kg. Se toma al azar una muestra de 55 estudiantes y se quiere saber la probabilidad de que la media de la muestra sea inferior a 6 kg. Esta probabilidad:

A) Se puede calcular con los datos del problema.

B) Se podría calcular con los datos del problema si nos aseguraran que los pesos de las mochilas se distribuyen según una distribución normal.

C) No se puede calcular sin saber el número total de los alumnos de instituto.

D) No se puede calcular, ya que el tamaño de la muestra no es lo suficientemente grande.

E) Se podría calcular si conociéramos la desviación típica de los pesos de la población de partida.

La respuesta correcta es la E, ya que la distribución en el muestreo de la media sigue una normal $N\left(\mu ; \frac{\sigma}{\sqrt{n}}\right)$.