12 Distribuciones muestrales

EJERCICIOS PROPUESTOS

1. Ejercicio resuelto.

2. El aeropuerto de una gran ciudad desea conocer la opinión de los viajeros acerca de los servicios que ofrece el aeropuerto.

a) Describe la población objetivo.

b) ¿Cómo elegirías una muestra representativa de la población objetivo?

a) Viajeros de entrada y/o salida del aeropuerto.

b) Para elegir una muestra representativa, debería realizarse un muestreo aleatorio estratificando al menos por días de la semana, horario (mañana, tarde y noche) y tipo de vuelo (nacional, internacional). Además, se puede estratificar por tramos de edad y por sexo. Debe tenerse en cuenta que, a mayor estratificación, se necesita mayor tamaño de la muestra total. En cambio, si no se estratifica convenientemente, no se pueden extraer conclusiones adecuadas para los grupos de interés.

3. Para realizar una encuesta de satisfacción con el gobierno municipal de una ciudad, se dispone del listado por orden alfabético de calles y del nombre de las personas que viven en cada domicilio. ¿Qué utilizarías para obtener una muestra, una lista con las direcciones o una lista con los nombres de las personas?

Para facilitar la recogida de información, se podría diseñar la obtención de la muestra en tres etapas:

i) En la primera etapa, se divide la ciudad en secciones homogéneas y se clasifican, por ejemplo, por tamaño de la población en cada sección, de mayor a menor. Se eligen al azar un determinado número de secciones.

ii) En la segunda etapa, se eligen aleatoriamente un determinado número de hogares (direcciones).

iii) En la tercera etapa, se sortea al azar una persona adulta de las que viven en el hogar.

El número de secciones y de domicilios de cada sección depende de la precisión con la que se deseen estimar los parámetros poblacionales y, claro está, del presupuesto disponible para llevar a cabo la encuesta.

4. La tercera parte de los alumnos de un centro estudia el idioma *A*, la mitad, el *B*, y el resto, el *C*. Cada alumno estudia solo uno de estos idiomas. Se selecciona una muestra de 60 alumnos, mediante muestreo aleatorio estratificado con asignación proporcional al número de alumnos de cada idioma.

a) ¿Cómo debería estar conformada la muestra?

b) En otra muestra seleccionada por el procedimiento anterior, el número de alumnos seleccionados del idioma *A* es 14.

Determina cuántos se han elegido de los otros dos idiomas.

a) La muestra debería elegirse proporcionalmente al número de alumnos que estudia cada uno de los idiomas.

Por tanto, la proporción de alumnos que estudian el idioma *C* es $1-\frac{1}{2}-\frac{1}{3}=\frac{1}{6}$.

N.º de alumnos seleccionados de cada idioma *A*: $\frac{1}{3}\cdot 60=20$, *B*: $\frac{1}{2}\cdot 60=30$ y *C*: $\frac{1}{6}\cdot 60=10$.

b) Sabemos que 14 es el número de alumnos seleccionados que estudian el idioma *A*, siendo estos un tercio del total. Si consideramos que *n* es el número de alumnos de la muestra, tenemos que $\frac{1}{3}n=14 \Rightarrow n=42$.

Como los alumnos del idioma *B* son la mitad, estos deben ser 21.

Por último, los alumnos que estudian el idioma *C* son 7, los que faltan para completar la muestra de 42 alumnos de la muestra.

5. Ejercicio resuelto.

6. La duración de un tipo de baterías para dispositivos móviles tiene una distribución normal de media 9 horas y de desviación típica 1,2 horas. Se toma una muestra aleatoria de 80 baterías. ¿Cuál es la probabilidad de que la duración media supere las 9,2 horas?

La variable X: "duración de la batería para dispositivos móviles" tiene una distribución $X \sim N(\mu = 9;\ \sigma = 1{,}2)$.

Si la muestra es de 80 baterías, la media muestral tiene una distribución normal:

$$\bar{X} \sim N\left(\mu = 9;\ \sigma^2 = \frac{1{,}2^2}{80}\right) = N\left(\mu = 9;\ \sigma^2 = 0{,}018\right)$$

Por tanto,

$$P\left(\bar{X} > 9{,}2\right) = P\left(Z > \frac{9{,}2 - 9}{\sqrt{0{,}018}}\right) = P\left(Z > 1{,}49\right) = 1 - P\left(Z < 1{,}49\right) = 1 - 0{,}9319 = 0{,}0681$$

7. En un aeropuerto, el tiempo de espera tiene distribución normal de media de 23 minutos con una desviación típica de 7 minutos. Si un viajero parte de dicho aeropuerto y viaja de lunes a viernes, calcular la probabilidad de que el tiempo medio de las 5 esperas sea superior a 25 minutos.

La media muestral de los 5 días, es una variable aleatoria $\bar{X} = \frac{1}{5}\sum_{i=1}^{5} X_i$ con distribución:

$$\bar{X} \sim N\left(\mu = 23;\ \sigma^2 = \frac{49}{5}\right) = N\left(\mu = 23;\ \sigma = \frac{7}{\sqrt{5}} = 3{,}1305\right)$$

Entonces:

$$P\left(\bar{X} > 25\right) = P\left(Z > \frac{25 - 23}{3{,}1305}\right) = P\left(Z > 0{,}64\right) = 1 - P\left(Z < 0{,}64\right) = 1 - 0{,}7389 = 0{,}2611$$

8. Ejercicio resuelto.

9. En las últimas elecciones el partido *A* ha recibido el 24 % de los votos emitidos. Si se elige una muestra aleatoria de 80 personas que han votado, se pide:

a) Distribución aproximada de la proporción de votantes al partido *A*.

b) ¿Cuál es la probabilidad de que hayan votado al partido *A* más de 25 personas de la muestra?

a) Sea la variable aleatoria X: "número de votantes del partido *A*, entre los 80". $X \sim \text{Bin}(n = 80;\ p = 0{,}24)$.

Sea $\hat{p}$ el estimador de la proporción de votantes del partido *A*. Como la muestra es suficientemente grande, la distribución de $\hat{p}$ se puede aproximar por

$$\hat{p} \sim N\left(\mu = 0{,}24;\ \sigma^2 = \frac{0{,}24(1 - 0{,}24)}{80}\right) = N\left(\mu = 0{,}24;\ \sigma^2 = 0{,}002\,28\right)$$

b) La probabilidad de que al partido *A* le hayan votado más de 25 personas en la muestra, puede calcularse aproximando la distribución de X por la de una variable aleatoria Y con distribución normal de media y varianza:

$$\mu = 80 \cdot 0{,}24 = 19{,}2 \qquad \sigma^2 = 80 \cdot 0{,}24 \cdot 0{,}76 = 14{,}592$$

Entonces, teniendo en cuenta la corrección por continuidad:

$$P\left(X > 25\right) \simeq P\left(Y \geq 25{,}5\right) = P\left(Z \geq \frac{25{,}5 - 19{,}2}{\sqrt{14{,}592}}\right) = P\left(Z \geq 1{,}65\right) = 1 - P\left(Z < 1{,}65\right) = 1 - 0{,}9505 = 0{,}0495$$

10. De una muestra aleatoria de 500 personas residentes en una determinada ciudad, 350 utilizan habitualmente el transporte público. Se pide:

a) Distribución aproximada de la proporción de usuarios habituales de transporte público.

b) Calcula la probabilidad de que menos de la cuarta parte de los habitantes no utilice el transporte público.

a) Si de la muestra aleatoria de 500 personas, 350 declararon utilizar habitualmente el transporte público, un estimador de la proporción es:

$$\hat{p} = \frac{350}{500} = 0,7$$

Dado que el tamaño de la muestra es suficientemente grande, la distribución de la proporción de usuarios habituales del transporte público es aproximadamente normal:

$$\hat{p} \sim N\left(\mu = p;\ \sigma^2 = \frac{p(1-p)}{n}\right) = N\left(\mu = 0,7;\ \sigma^2 = \frac{0,7(1-0,7)}{500} = 0,00042\right)$$

b) Utilizando la distribución aproximada de $\hat{p}$ obtenida en el apartado a), y teniendo en cuenta que el hecho de que menos de la cuarta parte (25 %) no utilice el trasporte público equivale a decir que más de las tres cuartas partes (75 %) sí lo utiliza, se tiene que:

$$P(\hat{p} > 0,75) = P\left(Z > \frac{0,75-0,7}{\sqrt{0,00042}}\right) = P(Z > 2,44) = 1 - P(Z < 2,44) = 1 - 0,9927 = 0,0073$$

11. Ejercicio resuelto.

12. El tiempo medio de atención a los pacientes en un centro médico es de 15 minutos, con una desviación típica de 8 minutos. Si en un determinado día se han citado 40 pacientes, calcula la probabilidad de que:

a) Se tarde más de 9 horas en atenderlos.

b) Se tarde en atenderlos menos de diez horas y media.

c) Se tarde entre 9 y 11 horas en atenderlos.

a) Sea X: "tiempo de atención a los pacientes del centro médico". X es una variable aleatoria, con media $\mu = 15$ minutos y desviación típica $\sigma = 8$ minutos, es decir, $X \sim N(\mu = 15;\ \sigma = 8)$.

Suponiendo que los 40 pacientes se han elegido al azar, la distribución de la suma de los tiempos de atención de atención a los 40 pacientes puede aproximarse por una distribución normal:

$$T_{40} \sim N\left(\mu_{40} = 40 \cdot 15 = 600;\ \sigma_{40}^2 = 40 \cdot 8^2 = 2560\right)$$

La probabilidad de que se tarde más de 9 horas (540 minutos) en atenderlos es:

$$P(T_{40} > 540) = P\left(Z > \frac{540-600}{\sqrt{2560}}\right) = P(Z > -1,19) = P(Z < 1,19) = 0,8830$$

b) La probabilidad de que tarde menos de 10 horas y media (630 minutos) en atenderlos es:

$$P(T_{40} < 630) = P\left(Z < \frac{630-600}{\sqrt{2560}}\right) = P(Z < 0,59) = 0,7224$$

c) La probabilidad de que tarde entre 9 horas (540minutos) y 11 horas (660 minutos) en atenderlos es:

$$P(540 \le T_{40} \le 660) = P\left(\frac{540-600}{\sqrt{2560}} \le Z \le \frac{660-600}{\sqrt{2560}}\right) = P(-1,19 \le Z \le 1,19) = P(Z \le 1,19) - P(Z \le -1,19) =$$

$$= P(Z \le 1,19) - (1 - P(Z \le 1,19)) = 2P(Z \le 1,19) - 1 = 2 \cdot 0,8830 - 1 = 0,7660$$

13. En una almazara se embotella el aceite en garrafas de 5 L. Después se agrupan en cajas que contienen 4 garrafas cada una. La máquina envasadora llena las garrafas según una distribución normal de media 5 L y desviación típica 5 cL.

a) Si un cliente compra una caja, calcula la probabilidad de que se lleve menos de 19,9 L de aceite.

b) Si un cliente compra 10 cajas, calcula la probabilidad de que se lleve menos de 199 L de aceite.

a) Se considera la variable *X*: “cantidad de aceite envasado en las garrafas de 5 L”, cuya distribución es normal $N(\mu = 5;\ \sigma = 0{,}05)$, en litros.

Cada caja contiene 4 garrafas, por lo que la suma de las 4 garrafas tiene una distribución normal de media $\mu_4 = 4 \cdot 5 = 20$ L y varianza $\sigma_4^2 = 4 \cdot 0{,}05^2 = 0{,}01$ L, es decir, $T_4 \sim N(\mu_4 = 20;\ \sigma_4^2 = 0{,}01)$.

Entonces, la probabilidad de que la caja lleve menos de 19,9 L de aceite es:

$$P(T_4 < 19{,}9) = P\left(Z < \frac{19{,}9 - 20}{\sqrt{0{,}01}}\right) = P(Z < -1) = 1 - P(Z < 1) = 1 - 0{,}8413 = 0{,}1587$$

b) Si un cliente compra 10 cajas, la suma de las 40 garrafas tiene una distribución normal de media $\mu_{40} = 40 \cdot 5 = 200$ L y varianza $\sigma_4^2 = 40 \cdot 0{,}05^2 = 0{,}1$ L, es decir, $T_{40} \sim N(\mu_{40} = 200;\ \sigma_{40}^2 = 0{,}1)$.

De manera que la probabilidad de que las 10 cajas lleven menos de 199 L es:

$$P(T_{40} < 199) = P\left(Z < \frac{199 - 200}{\sqrt{0{,}1}}\right) = P(Z < -3{,}16) = 1 - P(Z < 3{,}16) = 1 - 0{,}9992 = 0{,}0008$$

14. Ejercicio resuelto.

15. A una estación llegan autobuses procedentes de los puntos *A* y *B*. El tiempo medio de retraso del autobús procedente de *A* es de 10 minutos con una deviación típica de 4 minutos y el del procedente de *B* es de 15 minutos con una desviación típica de 5 minutos. Calcula la probabilidad de que el retraso medio de una muestra aleatoria de 25 días del autobús *B* no supere en 4 minutos el retraso medio de una muestra aleatoria de 30 días del autobús *A*.

Se consideran las variables aleatorias:

X: “tiempo de retraso del autobús procedente de *A*”, con media $\mu_X = 10$ y desviación típica $\sigma_X = 4$, en minutos.

Y: “tiempo de retraso del autobús procedente de *B*”, con media $\mu_Y = 15$ y desviación típica $\sigma_Y = 5$, en minutos.

Se eligen aleatoriamente $n_1 = 30$ días del autobús *A* y $n_2 = 25$ días del autobús *B*.

Dado que los tamaños muestrales son suficientemente grandes, las distribuciones de las medias muestrales de *X* e *Y* son, respectivamente, normales con la media y varianza que se indican:

$$\overline{X} \sim N\left(\mu_{\overline{X}} = 10;\ \sigma_{\overline{X}}^2 = \frac{4^2}{30} = 0{,}53\right) \qquad \overline{Y} \sim N\left(\mu_{\overline{Y}} = 15;\ \sigma_{\overline{Y}}^2 = \frac{5^2}{25} = 1\right)$$

La distribución de la diferencia de las medias muestrales del autobús procedente de *B* menos la del autobús procedente de *A* tiene media y varianza respectivas:

$$\mu_{\overline{Y}-\overline{X}} = \mu_{\overline{Y}} - \mu_{\overline{X}} = 15 - 10 = 5 \qquad \sigma_{\overline{Y}-\overline{X}}^2 = \sigma_{\overline{X}}^2 + \sigma_{\overline{Y}}^2 = 0{,}53 + 1 = 1{,}53$$

Por tanto, resulta: $\overline{Y} - \overline{X} \sim N(\mu_{\overline{Y}-\overline{X}} = 5;\ \sigma_{\overline{Y}-\overline{X}}^2 = 1{,}53)$, en minutos.

Entonces, la probabilidad de que el retraso medio de un autobús de *B* no supere en 4 minutos el retraso medio de un autobús de *A* es:

$$P(\overline{Y} - \overline{X} \le 4) = P\left(Z \le \frac{4 - 5}{\sqrt{1{,}53}}\right) = P(Z \le -0{,}81) = 1 - P(Z < 0{,}81) = 1 - 0{,}7910 = 0{,}2090$$

16. El peso medio de los pollos de una granja *A* es 2 kg con una desviación típica de 300 g. En otra granja *B* el peso medio es de 2,4 kg con una desviación típica de 400 g. Se eligen al azar 100 pollos de la granja *A* y 50 de la granja *B*. Calcula la probabilidad de que la suma de los pesos medios de ambas muestras sea mayor que 4,5 kg.

Se consideran las variables aleatorias:

X: "peso de los pollos de la granja *A*", con distribución de media $\mu_X = 2$ kg y desviación típica $\sigma_X = 0,3$ kg.

Y: "peso de los pollos de la granja *B*", con distribución de media $\mu_Y = 2,4$ kg y desviación típica $\sigma_Y = 0,4$ kg.

Se eligen aleatoriamente $n_1 = 100$ pollos de A y $n_2 = 50$ pollos de B.

Dado que los tamaños muestrales son suficientemente grandes, las distribuciones de las medias muestrales de *X* e *Y* son respectivamente, normales con la media y varianza que se indican:

$$\overline{X} \sim N\left(\mu_{\overline{X}} = 2;\ \sigma^2_{\overline{X}} = \frac{0,3^2}{100} = 0,0009\right) \qquad \overline{Y} \sim N\left(\mu_{\overline{Y}} = 2,4;\ \sigma^2_{\overline{Y}} = \frac{0,4^2}{50} = 0,0032\right)$$

Entonces, la distribución de la suma de las medias muestrales de la granja *A* y de la granja *B* es normal con media y varianza:

$$\mu_{\overline{X}+\overline{Y}} = \mu_{\overline{X}} + \mu_{\overline{Y}} = 2 + 2,4 = 4,4 \text{ kg} \qquad \sigma^2_{\overline{X}+\overline{Y}} = \sigma^2_{\overline{X}} + \sigma^2_{\overline{Y}} = 0,0009 + 0,0032 = 0,0041 \text{ kg}$$

De modo que la probabilidad de que la suma de los pesos medios sea superior a 4,5 kg es:

$$P\left(\overline{X} + \overline{Y} > 4,5\right) = P\left(Z > \frac{4,5 - 4,4}{\sqrt{0,0041}}\right) = P\left(Z > 1,56\right) = 1 - P\left(Z < 1,56\right) = 1 - 0,9406 = 0,0594$$

17. Ejercicio interactivo.

18 y 19. Ejercicios resueltos.

20. La proporción de artículos defectuosos en un gran lote de productos es 0,12. ¿Cuál debe ser el menor tamaño de la muestra aleatoria elegida para que, con una probabilidad de al menos 0,99, la proporción muestral de defectuosos sea menor que 0,15?

Si la muestra aleatoria de artículos tiene tamaño *n*, sea $\hat{p}$ el estimador de la proporción de artículos defectuosos. Suponiendo que *n* es suficientemente grande, por el teorema central de límite, la distribución de la proporción muestral se puede aproximar a la de una normal:

$$\hat{p} \sim N\left(\mu = 0,12;\ \sigma^2 = \frac{0,12(1 - 0,12)}{n} = \frac{0,1056}{n}\right)$$

Se exige que:

$$P\left(\hat{p} < 0,15\right) = 0,99$$

Tipificando y buscando en las tablas de la normal estándar:

$$P\left(Z < \frac{0,15 - 0,12}{\sqrt{0,1056}}\sqrt{n}\right) = 0,99 \Rightarrow P\left(Z < 0,0923\sqrt{n}\right) = 0,99 \Rightarrow 0,0923\sqrt{n} = 2,33 \Rightarrow n = \left(\frac{2,33}{0,0923}\right)^2 = 637,25$$

Por tanto, con un tamaño muestral de $n = 638$ se consigue el objetivo pedido.

21. De una población de 50 000 personas se estima que 1500 ven un determinado programa de televisión. Se selecciona una muestra aleatoria de 200 personas. ¿Cuál es la probabilidad aproximada de que menos de 4 personas hayan visto el programa?

La variable $X:\begin{cases} 1 & \text{si la persona seleccionada ve el programa} \\ 0 & \text{en otro caso} \end{cases}$, es una variable de Bernoulli con esperanza y varianza:

$$E[X] = \frac{1500}{50\,000} = 0{,}03 \qquad \mathrm{Var}(X) = 0{,}03(1-0{,}03) = 0{,}0291$$

De tal manera que $X_1, X_2, \ldots, X_{200}$ es una muestra aleatoria de una variable de Bernoulli, $\mathrm{Ber}(p = 0{,}03)$. Por el teorema central de límite, la variable suma, que es binomial $Y = \sum_{j=1}^{200} X_j \sim \mathrm{Bin}(n = 200;\ p = 0{,}03)$, se puede aproximar por una distribución normal de media y varianza:

$$\mu_{200} = 200 \cdot 0{,}03 = 6 \qquad \sigma_{200}^2 = 200 \cdot 0{,}0291 = 5{,}82$$

Entonces, se tiene que $Y \simeq T_{200} \sim N(\mu_{200} = 6;\ \sigma_{200}^2 = 5{,}82)$. Por tanto, utilizando la corrección por continuidad, la probabilidad de que menos de 4 personas, de las 200, hayan visto el programa es, aproximadamente:

$$P(Y < 4) \simeq P(T_{200} \le 3{,}5) = P\left(Z < \frac{3{,}5-6}{\sqrt{5{,}82}}\right) = P(Z < -1{,}04) = 1 - P(Z < 1{,}04) = 1 - 0{,}8508 = 0{,}1492$$

22. La estatura de los estudiantes de bachillerato de un centro escolar tiene una distribución de media $\mu = 170$ cm y desviación típica $\sigma = 5$ cm. Si se seleccionan 40 estudiantes al azar:

a) Determina la distribución aproximada de la media muestral.

b) Calcula la probabilidad de que la estatura media de la muestra sea inferior a 168 cm.

c) Calcula la probabilidad de que la suma de las estaturas de los 40 alumnos de la muestra sea superior a 68,5 metros.

a) La variable *X*: "estatura de los estudiantes de bachillerato" es una variable continua cuya distribución tiene media $\mu = 170$ cm y desviación típica $\sigma = 5$ cm.

Se selecciona una muestra de tamaño $n = 40$ y, aplicando el teorema central del límite, se tiene que la distribución de probabilidad de la media muestral es aproximadamente normal:

$$\overline{X} \sim N\left(\mu = 170;\ \sigma^2 = \frac{25}{40} = 0{,}625\right)$$

b) La probabilidad de que la media muestral sea inferior a 168 cm, es:

$$P(\overline{X} < 168) = P\left(Z < \frac{168-170}{\sqrt{0{,}625}}\right) = P(Z < -2{,}53) = 1 - P(Z < 2{,}53) = 1 - 0{,}9943 = 0{,}0057$$

c) La distribución de la suma de las estaturas de la muestra es, también, aproximadamente normal, con media y varianza:

$$\mu_{40} = 40 \cdot 170 = 6800 \text{ cm} \qquad \sigma_{40}^2 = 40 \cdot 25 = 1000 \text{ cm}^2$$

$$T_{40} = \sum_{i=1}^{40} X_i \sim N(\mu_{40} = 6800;\ \sigma_{40}^2 = 1000)$$

Si se tiene en cuenta que 68,5 m son 6850 cm, se tiene que:

$$P(T_{40} > 6850) = P\left(Z > \frac{6850-6800}{\sqrt{1000}}\right) = P(Z > 1{,}58) = 1 - P(Z < 1{,}58) = 1 - 0{,}9429 = 0{,}0571$$

23 a 29. Ejercicios resueltos.

EJERCICIOS

Población y muestra. Tipos de muestreo

30. Para estimar el número de erratas de un libro que contiene ilustraciones:

a) ¿Cuál será la población objetivo?

b) ¿Y la unidad de muestreo?

a) La población objetivo la constituyen las páginas que no contienen ilustraciones.

b) La unidad de muestreo es la página.

31. En las encuestas trimestrales del centro de investigaciones sociológicas (CIS) suele preguntarse, entre otras muchas cuestiones, por la puntuación que, de 1 a 10, le otorgan los ciudadanos a los líderes de los partidos políticos. La encuesta suele hacerse sobre 2500 personas.

a) Describe la población y la muestra.

b) ¿Qué técnica de muestreo será la más adecuada?

a) La población la constituyen los ciudadanos y ciudadanas mayores de edad con derecho a voto. La muestra estará formada por 2500 personas representativas de la población objeto de estudio.

b) Muestreo estratificado con al menos tres tipos de estratos: edades, tipo de ciudad (tamaño) y sexo. El procedimiento de muestreo se realiza en varias etapas, estratificando por conglomerados.

i) 1ª etapa: selección de los municipios.

ii) 2ª etapa: selección de las secciones dentro de los municipios de forma proporcional al tamaño de los municipios.

iii) 3ª etapa: selección de los individuos por rutas aleatorias y cuotas de sexo y edad.

Los estratos se forman por el cruce de las 17 comunidades autónomas con el tamaño de hábitat, dividido en 7 categorías: a) municipios de menos de 2001 habitantes; b) de 2001 a 10 000; c) de 10 001 a 50 000; d) de 50 001 a 100 000; e) de 100 001 a 400 000; f) de 400 001 a 1 000 000; y g) de más de 1 000 000 de habitantes.

Nota: Toda la información pertinente aparece en la página web del CIS.

32. Una universidad desea conocer el nivel de satisfacción de estudiantes, profesores y personal no docente. La universidad cuenta con dos campus y 22 facultades en total. Para ello encarga un estudio a una empresa consultora. Describe la población objetivo y la técnica de muestreo que sería más adecuada para obtener una muestra representativa de la población.

La población objetivo la forman todos los estudiantes matriculados en la universidad, así como los profesores y el personal no docente que trabajan en ella.

La técnica más adecuada sería un muestreo estratificado en dos etapas. En primer lugar se clasifican las facultades de cada campus por su tamaño y se selecciona una muestra de las mismas con afijación proporcional de acuerdo con su tamaño.

En la segunda etapa se procede a seleccionar, dentro de cada centro, una muestra aleatoria de cada colectivo, alumnos, profesores y personal no docente, proporcional al número de ellos en la facultad.

33. De los 15 000 empleados de una empresa multinacional, 9000 trabajan en labores de manufactura, 4500 en tareas administrativas o comerciales y el resto son cargos directivos. ¿Cómo se elegiría una muestra aleatoria de 250 personas representativa de los empleados de esta empresa?

La muestra de $n = 250$ empleados, debería elegirse mediante muestreo estratificado en tres estratos:

i) E1: empleados en labores de manufactura, con $N_1 = 9000$ personas.

ii) E2: empleados en tareas comerciales o administrativas, con $N_2 = 4500$ personas.

iii) E3: cargos directivos, con $N_3 = 1500$.

Dentro de cada estrato, el tamaño de la muestra, n_1, n_2 y n_3, debe ser proporcional al tamaño del estrato:

$$\frac{9000}{15\,000} = \frac{n_1}{250} \qquad \frac{4500}{15\,000} = \frac{n_2}{250} \qquad \frac{1500}{15\,000} = \frac{n_3}{250}$$

De esta forma, el tamaño de la muestra en cada estrato debe ser:

$$n_1 = \frac{9000 \cdot 250}{15\,000} = 150 \qquad n_2 = \frac{4500 \cdot 250}{15\,000} = 75 \qquad n_3 = \frac{1500 \cdot 250}{15\,000} = 25$$

Distribuciones muestrales de la media y la proporción

34. Una población tiene 5 elementos. Mediante muestreo aleatorio simple se seleccionan muestras de tamaño 3, siendo 2 la desviación típica de sus medias y 7 la media de las medias muestrales. ¿Cuánto valen la media y la varianza de la población?

Sea la variable X cuya media y varianza de la población son, respectivamente, μ y σ^2.

Si se extrae una muestra aleatoria de tamaño 3, la distribución de probabilidad de la media muestral $\overline{X}$ tiene media μ y varianza $\frac{\sigma^2}{3}$, es decir, $\mu = 7$ y $\frac{\sigma^2}{3} = 2^2 = 4$.

De manera que la media poblacional es $\mu = 7$ y la varianza poblacional es $\sigma^2 = 3 \cdot 4 = 12$.

35. Una población está formada por los elementos {2, 4, 5, 6}.

a) Escribe todas las muestras de tamaño 2 que pueden formarse mediante muestreo aleatorio simple con reemplazamiento y sin reemplazamiento.

b) Calcula para cada muestra la proporción de cifras impares.

c) Calcula la media y la varianza de la distribución muestral de la proporción de cifras impares.

a) Mediante muestreo con reemplazamiento, se pueden formar las siguientes muestras de tamaño 2:

$$R_1 = \{2,\ 2\} \quad R_3 = \{2,\ 5\} \quad R_5 = \{4,\ 4\} \quad R_7 = \{4,\ 6\} \quad R_9 = \{5,\ 6\}$$
$$R_2 = \{2,\ 4\} \quad R_4 = \{2,\ 6\} \quad R_6 = \{4,\ 5\} \quad R_8 = \{5,\ 5\} \quad R_{10} = \{6,\ 6\}$$

Y sin reemplazamiento: $M_1 = \{2,\ 4\}$ $M_2 = \{2,\ 5\}$ $M_3 = \{2,\ 6\}$ $M_4 = \{4,\ 5\}$ $M_5 = \{4,\ 6\}$ $M_6 = \{5,\ 6\}$

b) En cada caso, con y sin reemplazamiento, la proporción de cifras impares es:

Con reemplazamiento: $\hat{p}_{R1} = 0$ $\hat{p}_{R2} = 0$ $\hat{p}_{R3} = 0{,}5$ $\hat{p}_{R4} = 0$ $\hat{p}_{R5} = 0$ $\hat{p}_{R6} = 0{,}5$ $\hat{p}_{R7} = 0$ $\hat{p}_{R8} = 1$ $\hat{p}_{R9} = 0{,}5$ $\hat{p}_{R10} = 0$

Sin reemplazamiento: $\hat{p}_{M1} = 0$ $\hat{p}_{M2} = 0{,}5$ $\hat{p}_{M3} = 0$ $\hat{p}_{M4} = 0{,}5$ $\hat{p}_{M5} = 0$ $\hat{p}_{M6} = 0{,}5$

c) Se calcula la media y la varianza de la distribución muestral de la proporción de cifras impares:

Si es con reemplazamiento: $\overline{p}_R = \frac{1}{10}\sum_{i=1}^{10}\hat{p}_{Ri} = \frac{2{,}5}{10} = 0{,}25$ y $\text{Var}(\hat{p}_R) = \frac{1}{10}\sum_{i=1}^{10}\hat{p}_{Ri}^2 - \overline{p}_R^2 = \frac{1{,}75}{10} - 0{,}25^2 = 0{,}1125$

Si es sin reemplazamiento: $\overline{p}_M = \frac{1}{6}\sum_{j=1}^{6}\hat{p}_{Mj} = \frac{1{,}5}{6} = 0{,}25$ y $\text{Var}(\hat{p}_M) = \frac{1}{6}\sum_{i=1}^{6}\hat{p}_{Mi}^2 - \overline{p}_M^2 = \frac{0{,}75}{6} - 0{,}25^2 = 0{,}0625$

36. El 25 % de las personas que viven en una determinada ciudad son mayores de 65 años. De los habitantes de la ciudad se elige una muestra aleatoria de 250 personas.

a) Determina la distribución muestral de la proporción de personas mayores de 65 años.

b) Calcula la probabilidad de que más del 30% de la muestra sean personas mayores de 65 años.

c) Calcula la probabilidad de que el porcentaje de personas de la muestra sean menores de 65 años esté entre el 50 % y el 75 %.

a) La distribución de la proporción $\hat{p}$ de personas mayores de 65 años en la muestra de 250 personas es aproximadamente normal:

$$\hat{p} \sim N\left(\mu = 0{,}25;\ \sigma^2 = \frac{0{,}25(1-0{,}25)}{250} = 0{,}000\,75\right)$$

b) La probabilidad de que la proporción muestral de personas mayores de 65 años sea superior al 30 % se calcula utilizando la aproximación del aparatado anterior:

$$P(\hat{p} > 0{,}3) = P\left(Z > \frac{0{,}3-0{,}25}{\sqrt{0{,}000\,75}}\right) = P(Z > 1{,}83) = 1 - P(Z < 1{,}83) = 1 - 0{,}9664 = 0{,}0336$$

c) La distribución de la proporción $\hat{p}$ de personas menores de 65 años en la muestra de 250 personas es aproximadamente normal:

$$\hat{p} \sim N\left(\mu = 0{,}75;\ \sigma^2 = \frac{0{,}75(1-0{,}75)}{250} = 0{,}000\,75\right)$$

La probabilidad de que la proporción muestral de personas menores de 65 años esté entre el 50 % y el 75 % se calcula utilizando la aproximación anterior:

$$P(0{,}5 < \hat{p} < 0{,}75) = P\left(\frac{0{,}5-0{,}75}{\sqrt{0{,}000\,75}} < Z < \frac{0{,}75-0{,}75}{\sqrt{0{,}000\,75}}\right) = P(-9{,}13 < Z < 0) = P(Z < 0) - P(Z < -9{,}13) =$$
$$= 0{,}5 - 0 = 0{,}5.$$

37. Supongamos que el IMC (índice de masa corporal) en hombres adultos de una población sigue una distribución normal, $N(\mu,\ \sigma = 4)$.

a) Si el 6,68% de los hombres está en riesgo de sobrepeso, es decir, su IMC es superior a 25, calcula el valor del IMC medio, μ, para los hombres adultos de la población.

b) Si el IMC para los hombres adultos de la población sigue una distribución $N(\mu = 22;\ \sigma = 4)$ y se extrae una muestra aleatoria de 10 hombres adultos de esa población, calcula la probabilidad de que el IMC medio de la muestra esté entre los valores que la OMS considera normales, que son 18,5 y 25.

a) La variable aleatoria *X*: "IMC en hombres adultos" tiene distribución $N(\mu;\ \sigma = 4)$. Entonces, la información proporcionada afirma que $P(X > 25) = 0{,}0668$. Tipificando se obtiene:

$$P\left(Z > \frac{25-\mu}{4}\right) = 0{,}0668 \Rightarrow 1 - P\left(Z < \frac{25-\mu}{4}\right) = 0{,}0668 \Rightarrow P\left(Z < \frac{25-\mu}{4}\right) = 0{,}9332$$

Si se consulta en las tablas de la normal estándar se tiene que $\frac{25-\mu}{4} = 1{,}5 \Rightarrow \mu = 19$.

b) Si la variable *X* tiene una distribución $N(\mu = 22;\ \sigma = 4)$, la media muestral de una muestra de 64 hombres adultos también tiene distribución normal: $\bar{X} \sim N\left(\mu = 22;\ \sigma^2 = \frac{16}{10} = 1{,}6\right)$

De modo que la probabilidad de que el índice medio de la muestra esté entre 18,5 y 25 es:

$$P(18{,}5 < \bar{X} < 25) = P\left(\frac{18{,}5-22}{\sqrt{1{,}6}} Z < \frac{25-22}{\sqrt{1{,}6}}\right) = P(-2{,}77 < Z < 2{,}37) = P(Z < 2{,}37) - P(Z < -2{,}77) =$$
$$= P(Z < 2{,}37) - (1 - P(Z < 2{,}77)) = 0{,}9911 - 1 + 0{,}9972 = 0{,}9883$$

Distribución muestral de la suma

38. Una prueba consta de 100 preguntas tipo test, cada una con cuatro posibles respuestas de las cuales solo una es correcta. Si un estudiante contesta a las preguntas de forma aleatoria e independiente y para superar el examen es preciso responder correctamente al menos a 60 preguntas, ¿Cuál es la probabilidad de que apruebe?

La variable aleatoria $X:\begin{cases}0 & \text{si la respuesta no es correcta}\\ 1 & \text{si la respuesta sí es correcta}\end{cases}$, es una variable de Bernouilli $\text{Ber}(p=0{,}25)$ donde p = 0,25 es la probabilidad de acertar una respuesta si se responde al azar.

En la prueba se tienen 100 variables $X_1, X_2, \dots, X_{100}$ independientes e idénticamente distribuidas a X.

La suma de estas variables, la cual mide el número de respuestas acertadas, tiene una distribución binomial:

$$Y=\sum_{j=1}^{100} X_j \sim \text{Bin}(n=100;\ p=0{,}25)$$

Por el teorema central de límite, la distribución de Y se puede aproximar por la de una variable T con distribución normal:

$$T_{100} \sim N\left(\mu_{100}=100\cdot 0{,}25=25;\ \sigma_{100}^2=100\cdot 0{,}25\cdot 0{,}75=18{,}75\right)$$

Y, para calcular la probabilidad pedida, debe efectuarse la corrección por continuidad dado que la variable binomial es discreta y la normal es continua:

$$P(Y\ge 60)\simeq P(T_{100}\ge 59{,}5)=P\left(Z\ge\frac{59{,}5-25}{\sqrt{18{,}75}}\right)=P(Z\ge 7{,}97)=0$$

39. El tiempo de atención a un paciente, en una consulta médica, tiene una distribución normal de media 10 minutos y desviación típica igual a 3 minutos.

a) Si hay citados 8 pacientes, ¿cuál es la probabilidad de que sean atendidos en menos de 72 minutos?

b) Si hay citados 40 pacientes, ¿cuál es la probabilidad de que se necesiten más de 7 horas para atenderlos?

a) La variable X: "tiempo, en minutos, de atención al paciente" tiene una distribución $N(\mu=10;\ \sigma=3)$.

Si hay citados 8 pacientes, sus tiempos de atención representan una muestra aleatoria de tamaño $n=8$ de X y la suma tiene una distribución normal:

$$T_8=\sum_{j=1}^{8} X_j \sim N\left(\mu_8=8\cdot 10=80;\ \sigma_8^2=8\cdot 9=72\right)$$

De forma que la probabilidad de que los ocho pacientes sean atendidos en menos de 72 minutos es:

$$P(T_8<72)=P\left(Z<\frac{72-80}{\sqrt{72}}\right)=P(Z<-0{,}94)=1-P(Z<0{,}94)=1-0{,}8264=0{,}1736$$

b) Como en el apartado anterior, la suma de los tiempos de atención de los $n=40$ pacientes tiene una distribución normal:

$$T_{40}=\sum_{j=1}^{40} X_j \sim N\left(\mu_{40}=40\cdot 10=400;\ \sigma_{40}^2=40\cdot 9=360\right)$$

De tal manera que la probabilidad de que se necesiten más de 7 horas (420 minutos) para atender a los 40 pacientes es:

$$P(T_{40}>420)=P\left(Z>\frac{420-400}{\sqrt{360}}\right)=P(Z>1{,}05)=1-P(Z<1{,}05)=1-0{,}8531=0{,}1469$$

Distribuciones muestral de la suma y la diferencia de medias

40. La altura de los pinos (*pinus pinaster*) de una plantación *A* tiene una distribución con media 22 metros y desviación típica 1,5 metros. Mientras que los de otra plantación *B* tienen media 20 metros y desviación típica 1,25 metros. Se toman muestras aleatorias de 50 pinos de la plantación *A* y 25 de la *B*.

a) Escribe la distribución muestral de la media de cada una de las plantaciones.

b) Calcula la probabilidad de que la diferencia de las medias muestrales sea superior a un metro.

a) Se consideran las variables aleatorias:

X: "altura, en metros, de los pinos de la plantación *A*", con media $\mu_X = 22$ y varianza $\sigma_X^2 = 1{,}5^2 = 2{,}25$.

Y: "altura, en metros, de los pinos de la plantación *B*", con media $\mu_Y = 20$ y varianza $\sigma_Y^2 = 1{,}25^2 = 1{,}5625$.

La distribución de la media muestral de una muestra aleatoria de $n = 50$ pinos de la plantación *A*, se puede aproximar por la de una distribución normal $\overline{X} \sim N\left(\mu_{\overline{X}} = 22;\ \sigma_{\overline{X}}^2 = \dfrac{2{,}25}{50} = 0{,}045\right)$.

La distribución de la media muestral de una muestra aleatoria de $m = 25$ pinos de la plantación *B*, se puede aproximar por la de una distribución normal $\overline{Y} \sim N\left(\mu_{\overline{Y}} = 20;\ \sigma_{\overline{Y}}^2 = \dfrac{1{,}5625}{25} = 0{,}0625\right)$.

b) Se calcula primero la probabilidad de que la diferencia de las medias muestrales sea inferior a 1 metro, es decir, de que $\left|\overline{X} - \overline{Y}\right| < 1$. Como las muestras de *X* e *Y* son independientes, la diferencia de las medias muestrales tiene una distribución normal $\overline{X} - \overline{Y} \sim N\left(\mu_{\overline{X}-\overline{Y}} = 22 - 20 = 2;\ \sigma_{\overline{X}-\overline{Y}}^2 = 0{,}045 + 0{,}0625 = 0{,}1075\right)$.

De manera que:

$$P\left(\left|\overline{X} - \overline{Y}\right| < 1\right) = P\left(-1 < \overline{X} - \overline{Y} < 1\right) = P\left(\frac{-1-2}{\sqrt{0{,}1075}} < Z < \frac{1-2}{\sqrt{0{,}1075}}\right) = P\left(-9{,}15 < Z < -3{,}05\right) =$$

$$= P(Z < -3{,}05) - P(Z < -9{,}15) = \left(1 - P(Z < 3{,}05)\right) - P(Z < -9{,}15) = (1 - 0{,}9989) - 0 = 0{,}0011$$

Y, por tanto, $P\left(\left|\overline{X} - \overline{Y}\right| > 1\right) = 1 - P\left(\left|\overline{X} - \overline{Y}\right| < 1\right) = 1 - 0{,}0011 = 0{,}9989$.

Luego es prácticamente seguro que la diferencia de las medias muestrales sea superior a 1 metro.

41. En una población *A*, la tensión ocular de personas adultas tiene una distribución normal de media 18 y varianza 2, mientras que en otra población *B* tiene una distribución normal de media 20 y varianza 4. Se elige al azar una muestra de 25 personas adultas en cada una de las dos poblaciones.

a) Determina la distribución de la media muestral de cada población.

b) Escribe la distribución de la suma y de la diferencia de las medias muestrales.

c) Calcula la probabilidad de que la media muestral de la población *A* sea menor que la de la población *B* en al menos una unidad.

a) Se consideran las variables aleatorias:

X : "tensión ocular de las personas adultas de *A*", con distribución $X \sim N\left(\mu_X = 18;\ \sigma_X^2 = 2\right)$.

Y : "tensión ocular de las personas adultas de *B*", con distribución $Y \sim N\left(\mu_Y = 20;\ \sigma_Y^2 = 4\right)$.

La distribución de las medias muestrales de una muestra aleatoria de $n = 25$ personas adultas de cada una de las dos poblaciones tiene distribución normal:

$$\overline{X} \sim N\left(\mu_{\overline{X}} = 18;\ \sigma_{\overline{X}}^2 = \frac{2}{25} = 0{,}08\right) \qquad \overline{Y} \sim N\left(\mu_{\overline{Y}} = 20;\ \sigma_{\overline{Y}}^2 = \frac{4}{25} = 0{,}16\right)$$

b) Las distribuciones de la suma y de la diferencia de las medias muestrales son también normales:

$$\overline{X} + \overline{Y} \sim N\left(\mu_{\overline{X}+\overline{Y}} = 18 + 20 = 38;\ \sigma_{\overline{X}+\overline{Y}}^2 = 0{,}08 + 0{,}16 = 0{,}24\right)$$

$$\overline{X} - \overline{Y} \sim N\left(\mu_{\overline{X}-\overline{Y}} = 18 - 20 = -2;\ \sigma_{\overline{X}-\overline{Y}}^2 = 0{,}08 + 0{,}16 = 0{,}24\right)$$

c) Para este caso, la distribución de la diferencia de las medias muestrales es una normal:

$$\bar{Y}-\bar{X} \sim N\left(\mu_{\bar{Y}-\bar{X}}=20-18=2;\ \sigma^2_{\bar{Y}-\bar{X}}=0,08+0,16=0,24\right)$$

La probabilidad de que la media muestral de la población *A*, $\bar{X}$, sea menor que la de la población *B*, $\bar{Y}$, en al menos una unidad se calcula como sigue:

$$P\left(\bar{Y}-\bar{X}\geq 1\right)=P\left(Z\geq \frac{1-2}{\sqrt{0,24}}\right)=P\left(Z\geq -2,04\right)=P\left(Z<2,04\right)=0,9793$$

Teorema central del límite

42. Un servicio de urgencias recibe, en media, 8 avisos por hora con una varianza de 8. Si del servicio de urgencias se elige una muestra aleatoria de 150 horas:

a) Determina la distribución aproximada de la suma y de la media muestrales.

b) Calcula la probabilidad de que en la muestra se hayan recibido más de 1250 llamadas.

c) Calcula la probabilidad de que la media muestral esté entre 8 y 9 llamadas.

a) *X*: "número de avisos por hora en el servicio de urgencias", con distribución $X \sim N\left(\mu=8;\ \sigma^2=8\right)$.

Si se seleccionan al azar 150 horas de ese servicio de urgencias, las distribuciones de la suma del número de avisos en esas 150 horas y de la media muestral se pueden aproximar por las distribuciones de variables normales, dado que el tamaño muestral ($n=150$) es suficientemente grande. Así, se tiene que:

$$T_{150}=\sum_{j=1}^{150}X_j \sim N\left(\mu_{150}=150\cdot 8=1200;\ \sigma^2_{150}=150\cdot 8=1200\right) \qquad \text{y} \qquad \bar{X}\sim N\left(\mu_{\bar{X}}=8;\ \sigma^2_{\bar{X}}=\frac{8}{150}=0,0533\right)$$

b) Teniendo en cuenta el apartado anterior, queda que:

$$P\left(T_{150}>1250\right)=P\left(Z>\frac{1250-1200}{\sqrt{1200}}\right)=P\left(Z>1,44\right)=1-P\left(Z<1,44\right)=1-0,9251=0,0749$$

c) La probabilidad de que la media muestral esté entre 8 y 9 llamadas de urgencia es:

$$P\left(8\leq \bar{X}\leq 9\right)=P\left(\frac{8-8}{\sqrt{0,0533}}\leq Z\leq \frac{9-8}{\sqrt{0,0533}}\right)=P\left(0\leq Z\leq 4,33\right)=P\left(Z\leq 4,33\right)-P\left(Z\leq 0\right)=1-0,5=0,5$$

43. En una entidad bancaria, el 40% de los clientes tiene una hipoteca. Si se elige una muestra aleatoria de 50 clientes:

a) ¿Cuál es la distribución aproximada del número de clientes que tiene una hipoteca?

b) ¿Cuál es la distribución aproximada de la proporción de clientes que tienen una hipoteca?

c) Calcula la probabilidad de que más de 22 clientes de los 50 tengan una hipoteca.

a) La variable aleatoria $X:\begin{cases}0 & \text{si el cliente no tiene hipoteca}\\ 1 & \text{si el cliente tiene hipoteca}\end{cases}$, es una variable de Bernoulli $\text{Ber}\left(p=0,4\right)$, donde *p* es la probabilidad de que un cliente elegido al azar tenga hipoteca.

Si consideramos la muestra $X_1, X_2,\ldots, X_{50}$, la variable $Y=\sum_{j=1}^{50}X_j \sim \text{Bin}\left(n=50;\ p=0,4\right)$ se puede aproximar por una distribución normal $T_{50}\sim N\left(\mu_{50}=50\cdot 0,4=20;\ \sigma^2_{50}=50\cdot 0,4\cdot 0,6=12\right)$.

b) Dado que el tamaño muestral es suficientemente grande, la distribución de la proporción $\hat{p}$ se puede aproximar por la de una distribución normal $\hat{p}\sim N\left(\mu=0,4;\ \sigma^2=\frac{0,4\cdot 0,6}{50}=0,0048\right)$.

c) Se utiliza la aproximación normal teniendo la precaución de efectuar la corrección por continuidad, dado que la variable binomial es discreta y la normal es continua:

$$P\left(Y>22\right)\simeq P\left(T_{50}\geq 22,5\right)=P\left(Z\geq \frac{22,5-20}{\sqrt{12}}\right)=P\left(Z\geq 0,72\right)=1-P\left(Z<0,72\right)=1-0,7642=0,2358$$

CUESTIONES

44. Señala cuál sería la población objetivo e indica la conveniencia o no de extraer una muestra en cada uno de los casos.

a) La media de la estatura de los asistentes a una manifestación.

b) El tiempo de vida de un tipo de batería para móviles.

c) El número de clientes potenciales de un nuevo hipermercado.

a) La población objetivo son todos los asistentes a dicha manifestación. En principio, mientras la asistencia no fuera extremadamente elevada, se podría realizar un estudio individuo a individuo. Por tanto, no haría falta extraer ninguna muestra.

b) La población objetivo son todas las baterías de ese determinado tipo. En este caso, habría que extraer una muestra ya que realizar un estudio de cada elemento de la población objetivo, resultaría inviable debido a los costes que esto supondría; bien por ser la muestra extremadamente grande, o bien por otros factores tales como, por ejemplo, que ese modelo de batería se haya comercializado por todo el mundo.

c) La población objetivo son todos los individuos que pueden ser clientes del nuevo hipermercado (población que viva cerca del hipermercado, personas que estén interesadas en los tipos de productos que ofrezca dicho hipermercado, personas que encuentren interesantes los precios del hipermercado, etc.). En este caso, sí convendría extraer una muestra ya que resulta muy costoso tener acceso a todos los posibles clientes que puedan estar interesados en comprar en este hipermercado.

45. En los siguientes casos, ¿qué proceso de muestreo resultaría el más adecuado?

a) Se quiere estimar el porcentaje de ciudadanos con estudios superiores en una determinada ciudad seleccionando una muestra de hogares y realizando una encuesta por teléfono.

b) Para estimar los ingresos medios de la población de 18 a 65 años de edad de un país se selecciona una muestra.

a) Lo primero que se debe tener en cuenta es el número de habitantes que tiene dicha ciudad. Si éste fuera muy grande, obtener un listado de todos los ciudadanos para llevar a cabo un muestreo aleatorio simple y que todos tuvieran la misma probabilidad de ser incluidos en la muestra, resultaría poco viable puesto que, por ejemplo, pueden haber personas que estén censadas en dicha ciudad pero que no residan en ella. Por tanto, resulta más efectivo realizar un muestreo estratificado, estratificando la población, por ejemplo, por barrios o secciones censales.

b) Si el país no tuviera grandes diferencias de población entre sus distintas regiones, un muestreo por conglomerado o por estratos resultaría muy acertado. Sin embargo, si hubiese muchas diferencias de población entre unas regiones y otras, realizar un muestreo aleatorio simple a través de un listado de todas las personas que habitan en dicho país (entre 18 y 65 años) sería la forma más efectiva de realizar el proceso de muestreo.

PROBLEMAS

46. El 10 % de un gran lote de artículos manufacturados es defectuoso. Del lote de artículos se selecciona una muestra aleatoria de tamaño *n*.

a) Si n = 40, ¿cuál es la probabilidad de que en la muestra el porcentaje de artículos defectuosos sea superior al 9 %?

b) ¿Cuál debe ser el mínimo tamaño muestral para que con probabilidad 0,95, el porcentaje de artículos defectuosos en la muestra sea inferior al 13 %?

a) La variable aleatoria $X : \begin{cases} 0 & \text{si el artículo no es defectuoso} \\ 1 & \text{si el artículo es defectuoso} \end{cases}$ es una variable de Bernoulli $\text{Ber}(p = 0{,}1)$.

Se selecciona una muestra aleatoria de n artículos, que se corresponde con una muestra aleatoria de $X_1, X_2, \dots, X_n$ de la variable X. Así, si $n = 40$, la distribución de la proporción $\hat{p}$ se puede aproximar por una normal: $\hat{p} \sim N\left(\mu = 0{,}1;\ \sigma^2 = \dfrac{0{,}1 \cdot 0{,}9}{40} = 0{,}00225\right)$. Por tanto, queda que:

$$P(\hat{p} > 0{,}09) = P\left(Z > \frac{0{,}09 - 0{,}1}{\sqrt{0{,}00225}}\right) = P(Z > -0{,}21) = 1 - P(Z < 0{,}21) = 1 - 0{,}5832 = 0{,}4168$$

b) Se tiene que la distribución de la proporción es aproximadamente normal $\hat{p} \sim N\left(\mu = 0{,}1;\ \sigma^2 = \dfrac{0{,}09}{n}\right)$.

$$P(\hat{p} < 0{,}13) = 0{,}95 \Rightarrow P\left(Z < \frac{0{,}13 - 0{,}1}{\sqrt{0{,}09}}\sqrt{n}\right) = 0{,}95 \Rightarrow P\left(Z < 0{,}1\sqrt{n}\right) = 0{,}95 \Rightarrow 0{,}1\sqrt{n} = 1{,}645 \Rightarrow n = 270{,}6$$

Luego es necesario tener una muestra de 271 artículos.

47. En una fiesta se inflan globos de una bombona de helio de 40 L. Se espera que, en promedio, cada globo que se infle contenga 20 cL, con una desviación típica de 8 cL. Los globos se inflan de forma independiente. Calcula la probabilidad de que la bombona quede vacía después de inflar 190 globos.

La variable X: "contenido, en cL, de cada globo" sigue una distribución $X \sim N(\mu = 20;\ \sigma^2 = 8^2 = 64)$, en cL.

La variable aleatoria suma de los contenidos de la muestra de 190 globos tiene una distribución que puede aproximarse mediante la de una normal $T_{190} = \sum_{j=1}^{190} X_j \sim N(\mu_{190} = 190 \cdot 20 = 3800;\ \sigma_{190}^2 = 190 \cdot 64 = 12\,160)$.

La bombona quedará vacía si la suma del contenido de los 190 globos supera los 4000 cL (40 L), entonces:

$$P(T_{190} \geq 4000) = P\left(Z \geq \frac{4000 - 3800}{\sqrt{12160}}\right) = P(Z \geq 1{,}81) = 1 - P(Z < 1{,}81) = 1 - 0{,}9649 = 0{,}0351$$

48. El error en centigramos, que se comete al pesar un determinado objeto, es una variable aleatoria con distribución normal de media $\mu = 0$ y varianza $\sigma^2 = 100$. Se pesa el objeto *n* veces.

a) Determina la distribución del error medio de las n pesadas.

b) Si se pretende que con probabilidad 0,99 el error medio cometido no supere los 4 centigramos, ¿cuántas pesadas deben hacerse?

a) La variable aleatoria X: "error, en centigramos, al pesar un objeto" tiene distribución $X \sim N(\mu = 0;\ \sigma^2 = 100)$.

Se pesa n veces el objeto, es decir, se toma una muestra aleatoria de tamaño n.

La distribución del error medio es normal $\overline{X} \sim N\left(\mu = 0;\ \sigma^2 = \dfrac{100}{n}\right)$.

b) Si no se quiere que el error medio cometido supere los 4 centigramos, entonces:

$$P(\overline{X} \leq 4) = 0{,}99 \Rightarrow P\left(Z \leq \frac{4-0}{10}\sqrt{n}\right) = 0{,}99 \Rightarrow P\left(Z \leq 0{,}4\sqrt{n}\right) = 0{,}99 \Rightarrow 0{,}4\sqrt{n} = 2{,}33 \Rightarrow n = 33{,}9$$

Por tanto, se necesitan 34 pesadas.

49. En un laboratorio, una investigadora lleva a cabo sucesivas pesadas de un objeto con una báscula de precisión que, no obstante, en cada medición, tiene una desviación típica σ. La investigadora realiza 25 pesadas del objeto. Utiliza el teorema central del límite para hallar aproximadamente la probabilidad de que la investigadora estime el verdadero peso del objeto con una precisión superior a la cuarta parte de la desviación típica.

La variable X: "peso del objeto en la báscula", tiene media μ y desviación típica σ, ambas desconocidas.

Se realizan 25 pesadas del objeto, es decir, se considera una muestra aleatoria de X de tamaño n = 25.

La distribución de la media muestral se puede aproximar por una distribución normal: $\overline{X} \sim N\left(\mu;\ \dfrac{\sigma^2}{25}\right)$.

Si la media muestral debe estimar el valor de μ con precisión superior a un cuarto de la desviación típica σ, esto implica que:

$$P\left(\left|\overline{X} - \mu\right| < \frac{\sigma}{4}\right) = P\left(-\frac{\sigma}{4} < \overline{X} - \mu < \frac{\sigma}{4}\right) = P\left(-\frac{\sigma}{4}\cdot\frac{5}{\sigma} < Z < \frac{\sigma}{4}\cdot\frac{5}{\sigma}\right) = P\left(-\frac{5}{4} < Z < \frac{5}{4}\right) = P(-1{,}25 < Z < 1{,}25) =$$
$$= P(Z < 1{,}25) - P(Z < -1{,}25) = P(Z < 1{,}25) - (1 - P(Z < 1{,}25)) = 2P(Z < 1{,}25) - 1 = 2 \cdot 0{,}8944 - 1 = 0{,}7888$$

50. El cociente intelectual de los individuos presentes en una sala puede suponerse que sigue una distribución normal de media μ y varianza igual a 81.

a) ¿Cuánto vale μ si sabemos que solo un 10 % de las personas en la sala supera un cociente intelectual de 105?

En los dos siguientes apartados se considera que $\mu = 95$:

b) Elegida al azar una persona de la sala, ¿cuál es la probabilidad de que su cociente intelectual esté entre 86 y 107?

c) Elegimos 9 personas al azar de la sala y se considera la media de sus cocientes intelectuales. ¿Cuál es la probabilidad de que esa media esté entre 86 y 107?

a) La variable aleatoria *X*: "cociente intelectual de los individuos presentes en la sala" sigue $X \sim N\left(\mu;\ \sigma^2 = 81\right)$.

Si el 10 % supera el cociente intelectual 105, se expresa:

$$P(X > 105) = 0{,}1 \Rightarrow P\left(Z > \frac{105-\mu}{9}\right) = 0{,}1 \Rightarrow 1 - P\left(Z < \frac{105-\mu}{9}\right) = 0{,}1 \Rightarrow P\left(Z < \frac{105-\mu}{9}\right) = 0{,}9$$

Y, por tanto, usando la tabla de la normal estándar queda $\dfrac{105-\mu}{9} = 1{,}282 \Rightarrow \mu = 93{,}462.$

b) Si $X \sim N\left(\mu = 95;\ \sigma^2 = 81\right) \Rightarrow P(86 < X < 107) = P\left(\dfrac{86-95}{9} < Z < \dfrac{107-95}{9}\right) = P(-1 < Z < 1{,}33) =$

$$= P(Z < 1{,}33) - P(Z < -1) = P(Z < 1{,}33) - \left(1 - P(Z < 1)\right) = 0{,}9082 - 1 + 0{,}8413 = 0{,}7495$$

c) Con una muestra aleatoria de 9 personas, se tiene que la media muestral $\bar{X} \sim N\left(\mu = 95;\ \sigma^2 = \dfrac{81}{9} = 9\right)$.

$$P\left(86 < \bar{X} < 107\right) = P\left(\frac{86-95}{3} < Z < \frac{107-95}{3}\right) = P(-3 < Z < 4) = P(Z < 4) - P(Z < -3) =$$
$$= P(Z < 4) - \left(1 - P(Z < 3)\right) = 1 - 1 + 0{,}9987 = 0{,}9987$$

51. En una cabaña ganadera, el 20 % de las reses se ha visto afectada por una enfermedad no mortal. Si de esta cabaña se selecciona una muestra aleatoria de 80 reses:

a) ¿Cuál es la distribución de la proporción muestral?

b) Calcula la probabilidad de que la proporción muestral de reses que han estado enfermas sea inferior al 18 %.

c) Determina la distribución muestral de la suma de reses que estuvieron enfermas.

a) La variable $X: \begin{cases} 0 & \text{si la res no ha enfermado} \\ 1 & \text{si la res ha enfermado} \end{cases}$, es una variable aleatoria de Bernoulli $Ber\,(p = 0{,}2)$, siendo $p = 0{,}2$ la probabilidad de que una res elegida al azar haya enfermado. De la cabaña ganadera se elige una muestra aleatoria de 80 reses: $X_1, X_2, \ldots, X_{80}$.

La distribución de la proporción muestral, $\hat{p} = \dfrac{1}{80}\sum_{j=1}^{80} X_j$, se puede aproximar por la de una distribución normal, dado que el tamaño muestral es suficientemente grande: $\hat{p} \sim N\left(\mu = 0{,}2;\ \sigma^2 = \dfrac{0{,}2 \cdot 0{,}8}{80} = 0{,}002\right)$

b) Para calcular la probabilidad pedida se utiliza la aproximación del apartado anterior:

$$P(\hat{p} < 0{,}18) = P\left(Z < \frac{0{,}18 - 0{,}2}{\sqrt{0{,}002}}\right) = P(Z < -0{,}45) = 1 - P(Z < 0{,}45) = 1 - 0{,}6736 = 0{,}3264$$

c) La distribución muestral de la suma de reses que estuvieron enfermas se aproxima por una normal:

$$T_{80} = \sum_{j=1}^{80} X_j \sim N\left(\mu_{80} = 80 \cdot 0{,}2 = 16;\ \sigma_{80}^2 = 80 \cdot 0{,}002 = 0{,}16\right) \Leftrightarrow T_{80} \sim N\left(\mu_{80} = 16;\ \sigma_{80}^2 = 0{,}16\right)$$

52. En una determinada región, el salario medio de los hombres es de 1500 € al mes, con una desviación típica de 300 €, y el de las mujeres, de 1100 €, con una desviación típica de 200 €. Si de esta región, se elige una muestra aleatoria de 100 mujeres y 80 hombres:

a) Escribe las distribuciones de las sumas y las medias muestrales de hombres y mujeres.

b) Calcula la probabilidad de que la suma de las medias muestrales de hombres y mujeres supere los 2500 € al mes

c) Calcula la probabilidad de que la diferencia de las medias muestrales entre hombres y mujeres supere los 450 €.

a) Se consideran las variables aleatorias:

X: "salario mensual de los hombres", con media y desviación típica respectivas $\mu_X = 1500$ € y $\sigma_X = 300$ €.

Y: "salario mensual de las mujeres", con media y desviación típica respectivas $\mu_Y = 1100$ € y $\sigma_Y = 200$ €.

Se extraen muestras aleatorias independientes de los salarios de 80 hombres y 100 mujeres:

$$X_1, X_2, \ldots, X_{80} \quad \text{e} \quad Y_1, Y_2, \ldots, Y_{100}$$

Las distribuciones de las sumas y las medias muestrales son aproximadamente normales:

Para los hombres:

$$T_{80} = \sum_{i=1}^{80} X_i \sim N\left(\mu_{80} = 80 \cdot 1500 = 120\,000;\ \sigma_{80}^2 = 80 \cdot 300^2 = 7\,200\,000\right)$$

$$\overline{X} \sim N\left(\mu = 1500;\ \sigma^2 = \frac{300^2}{80} = 1125\right)$$

Para las mujeres:

$$T_{100} = \sum_{i=1}^{100} Y_i \sim N\left(\mu_{100} = 100 \cdot 1100 = 110\,000;\ \sigma_{100}^2 = 100 \cdot 200^2 = 4\,000\,000\right)$$

$$\overline{Y} \sim N\left(\mu = 1100;\ \sigma^2 = \frac{200^2}{100} = 400\right)$$

b) La distribución de la suma de las medias muestrales de hombres y mujeres se puede aproximar por la de una distribución normal como sigue:

$$\overline{X} + \overline{Y} \sim N\left(\mu_{\overline{X}+\overline{Y}} = 1500 + 1100 = 2600;\ \sigma_{\overline{X}+\overline{Y}}^2 = 1125 + 400 = 1525\right)$$

Por lo que la probabilidad que se pide se obtiene de la siguiente manera:

$$P\left(\overline{X} + \overline{Y} > 2500\right) = P\left(Z > \frac{2500 - 2600}{\sqrt{1525}}\right) = P(Z > -2{,}56) = P(Z < 2{,}56) = 0{,}9948$$

c) La distribución de la diferencia de las medias muestrales de hombres y mujeres se puede aproximar por la de una distribución normal como sigue:

$$\overline{X} - \overline{Y} \sim N\left(\mu_{\overline{X}-\overline{Y}} = 1500 - 1100 = 400;\ \sigma_{\overline{X}-\overline{Y}}^2 = 1125 + 400 = 1525\right)$$

La probabilidad que se pide, $P\left(\left|\overline{X} - \overline{Y}\right| > 450\right)$, se obtiene de la siguiente manera. Primero calculamos la probabilidad de que la diferencia no supere los 450 €:

$$P\left(\left|\overline{X} - \overline{Y}\right| < 450\right) = P\left(-450 < \overline{X} - \overline{Y} \le 450\right) = P\left(\frac{-450 - 400}{\sqrt{1525}} \le Z \le \frac{450 - 400}{\sqrt{1525}}\right) = P(-21{,}77 \le Z \le 1{,}28) =$$

$$= P(Z \le 1{,}28) - P(Z \le -21{,}77) = 0{,}8997 - 0 = 0{,}8997$$

Sabiendo esto, se tiene que:

$$P\left(\left|\overline{X} - \overline{Y}\right| > 450\right) = 1 - P\left(\left|\overline{X} - \overline{Y}\right| \le 450\right) = 1 - 0{,}8997 = 0{,}1003$$

53. La duración de una conocida marca de bombillas tiene una distribución normal de media 1380 horas y desviación típica 390 horas. Se extrae una muestra aleatoria de tamaño 10 que presentó una duración media de 1425 horas. Otra muestra aleatoria de tamaño 8, independiente de la anterior, presentó una duración media de 1342 horas. Calcula la probabilidad de que las medias muestrales presenten una diferencia mayor de la que se ha obtenido en esta ocasión.

Sea X: "duración, en horas, de las bombillas de esta marca". Se sabe que $X \sim N(\mu = 1380;\ \sigma = 390)$.

La media de la muestra de tamaño 10 tiene distribución normal:

$$\overline{X}_{10} \sim N\left(\mu = 1380;\ \sigma_{10}^2 = \frac{390^2}{10} = 15\,210\right)$$

La media de la muestra de tamaño 8 tiene distribución normal:

$$\overline{X}_{8} \sim N\left(\mu = 1380;\ \sigma_{8}^2 = \frac{390^2}{8} = 19\,012{,}5\right)$$

La distribución de la diferencia de las medias muestrales es $\overline{X}_{10} - \overline{X}_8 \sim N\left(\mu = 0;\ \sigma^2_{\overline{X}_{10}-\overline{X}_8} = 34\,222{,}5\right)$.

En esta ocasión la diferencia observada ha sido de 1425 – 1342 = 83 horas.

La probabilidad de que esa diferencia sea superior a 83 horas es $P\left(\left|\overline{X}_{10} - \overline{X}_8\right| > 83\right)$.

Primero se calcula la siguiente probabilidad:

$$P\left(\left|\overline{X}_{10} - \overline{X}_8\right| \le 83\right) = P\left(-83 \le \overline{X}_{10} - \overline{X}_8 \le 83\right) = P\left(\frac{-83}{\sqrt{34\,222{,}5}} \le Z \le \frac{83}{\sqrt{34\,222{,}5}}\right) = P(-0{,}45 \le Z \le 0{,}45) =$$

$$= P(Z \le 0{,}45) - P(Z \le -0{,}45) = P(Z \le 0{,}45) - (1 - P(Z \le 0{,}45)) = 2P(Z \le 0{,}45) - 1 = 2 \cdot 0{,}6736 - 1 = 0{,}3472$$

Así, se obtiene que $P\left(\left|\overline{X}_{10} - \overline{X}_8\right| > 83\right) = 1 - P\left(\left|\overline{X}_{10} - \overline{X}_8\right| \le 83\right) = 1 - 0{,}3472 = 0{,}6528$.

AUTOEVALUACIÓN

Comprueba qué has aprendido

1. En un centro donde se estudian idiomas, 440 alumnos estudian inglés, 122, alemán, y 80, otros idiomas. Se supone que un alumno solo cursa uno de los idiomas. Si se quiere conocer el nivel de satisfacción de los estudiantes matriculados en el centro ¿cómo debería elegirse una muestra de 120 alumnos, que sea representativa del conjunto de alumnos del centro?

Debe elegirse una muestra aleatoria estratificada, con afijación proporcional al tamaño de cada estrato.

Se tienen tres estratos:

i) E1: "alumnos que estudian inglés", con $N_1 = 440$ alumnos.

ii) E2: "alumnos que estudian alemán" con $N_2 = 122$ alumnos. Total: 642 alumnos

iii) E3: "alumnos que estudian otros idiomas" con $N_3 = 80$ alumnos.

Para elegir el tamaño muestral de cada estrato, n_1, n_2 y n_3, se reparten los 120 alumnos, que deben formar parte de la muestra, proporcionalmente al tamaño de cada estrato:

$$\frac{120}{642} = \frac{n_1}{440} = \frac{n_2}{122} = \frac{n_3}{80}$$

Con lo que

$$n_1 = \frac{120 \cdot 440}{642} = 82{,}24;\ n_2 = \frac{120 \cdot 122}{642} = 22{,}8;\ n_3 = \frac{120 \cdot 80}{642} = 14{,}95$$

Por tanto, la muestra debería elegirse con 82 alumnos de inglés, 23 de alemán y 15 de otros idiomas.

2. En un conocido concurso de televisión el premio que se lleva el ganador se va acumulando si no se completa la prueba final, y aumenta cada programa una media de 2558 € con una desviación típica de 280 €. Si se elige una muestra aleatoria de 40 programas emitidos:

a) ¿Cuál será la distribución de la suma de los premios acumulados en los 40 programas?

b) Halla la probabilidad de que dicha suma supere los 100 000 €.

a) Sea X: "cantidad, en euros, que aumenta el bote cada programa". La distribución de X tiene media $\mu = 2558$ y desviación típica $\sigma = 280$, en euros.

Se elige una muestra aleatoria de 40 programas: $X_1, X_2, \ldots, X_{40}$.

Como el tamaño de la muestra es suficientemente grande, la distribución de la suma de los 40 premios se puede aproximar por la de una distribución normal:

$$T_{40} = \sum_{j=1}^{40} X_j \sim N\left(\mu_{40} = 40 \cdot 2558 = 102\,320;\ \sigma_{40}^2 = 40 \cdot 280^2 = 3\,136\,000\right)$$

b) La probabilidad de que la suma supere los cien mil euros:

$$P\left(T_{40} > 100\,000\right) = P\left(Z > \frac{100\,000 - 102\,320}{\sqrt{3\,136\,000}}\right) = P\left(Z > -1{,}31\right) = P\left(Z < 1{,}31\right) = 0{,}9049$$

3. Una epidemia de gripe ha afectado al 60 % de la plantilla de una gran empresa. Si se elige una muestra aleatoria de 125 empleados, se pide:

a) Distribución de la proporción muestral de afectados.

b) Calcula la probabilidad de que la gripe haya afectado a más de 70 empleados de la muestra.

a) Se considera la variable aleatoria X: "la gripe ha afectado al empleado", que es una variable aleatoria de Bernoulli con $p = 0{,}6$, $\text{Ber}\left(p = 0{,}6\right)$. Se elige una muestra aleatoria de 125 empleados y se observa las variables $X_1, X_2, \ldots, X_{125}$ independientes e idénticamente distribuidas que X.

La distribución de la proporción del número de afectados, se puede aproximar por la distribución normal de media $\mu = 0{,}6$ y varianza $\sigma^2 = \dfrac{0{,}6\left(1 - 0{,}6\right)}{125} = 0{,}00192$. Así, $\hat{p} \sim N\left(\mu = 0{,}6;\ \sigma^2 = 0{,}00192\right)$.

b) 70 empleados de 125, representa el 56% de los empleados de la muestra. Por tanto:

$$P\left(\hat{p} > 0{,}56\right) = P\left(Z > \frac{0{,}56 - 0{,}6}{\sqrt{0{,}00192}}\right) = P\left(Z > -0{,}91\right) = P\left(Z < 0{,}91\right) = 0{,}8186$$

4. El peso de los bebés al nacer en un hospital tiene una distribución normal de media 3,25 kg y desviación típica 500 g. Se elige una muestra aleatoria de 20 bebés nacidos en dicho hospital.

a) Calcula la distribución de la media muestral del peso de los bebés.

b) Calcula la probabilidad de que el peso medio supere los 3,5 kg.

a) La variable X: "peso en kg de los bebés al nacer" tiene una distribución $X \sim N\left(\mu = 3{,}25; \sigma = 0{,}5\right)$, en kilos.

Se elige una muestra aleatoria $X_1, X_2, \ldots, X_{20}$. La distribución de la media muestral es también normal:

$$\bar{X} \sim N\left(\mu = 3{,}25;\ \sigma^2 = \frac{0{,}5^2}{20} = 0{,}0125\right)$$

b) Entonces, la probabilidad que se pide es:

$$P\left(\bar{X} > 3{,}5\right) = P\left(Z > \frac{3{,}5 - 3{,}25}{\sqrt{0{,}0125}}\right) = P\left(Z > 2{,}24\right) = 1 - P\left(Z < 2{,}24\right) = 1 - 0{,}9875 = 0{,}0125$$

Relaciona y contesta

Elige la única respuesta correcta en cada caso

1. Para acceder a una determinada universidad, se deben superar unas pruebas tipo test con respuesta múltiple. En la última convocatoria, la proporción de aciertos fue del 70 %. Si se elige una muestra aleatoria de 50 estudiantes, la proporción muestral de aciertos tiene una distribución:

A. $N\left(\mu = 0{,}7;\ \sigma^2 = 0{,}065\right)$

B. $N\left(\mu = 0{,}7;\ \sigma^2 = 0{,}0042\right)$

C. $N\left(\mu = 0{,}3;\ \sigma^2 = 0{,}0042\right)$

D. $N\left(\mu = 0{,}3;\ \sigma^2 = 0{,}065\right)$

La solución es B. La media es $\mu = \frac{70}{100} = 0{,}7$ y $\sigma^2 = \frac{0{,}7\cdot(1-0{,}7)}{50} = 0{,}0042$.

2. Las notas de matemáticas en una clase de 2.º de Bachillerato tienen una distribución normal de media 6 y desviación típica 2. Se elige una muestra de 6 alumnos. La probabilidad de que la media de la muestra sea superior a 7 puntos es:

A. 0,1112 **B.** 0,8888 **C.** 0,9987 **D.** 0,0013

La solución es A. Tenemos que $\bar{X} \sim N\left(\mu = 6;\ \sigma^2 = \frac{2^2}{6} = 0{,}67\right)$. Por tanto, queda que:

$$P\left(\bar{X} > 7\right) = P\left(Z > \frac{7-6}{\sqrt{0{,}67}}\right) = P\left(\bar{X} > 1{,}22\right) = 1 - P\left(\bar{X} < 1{,}22\right) = 1 - 0{,}8888 = 0{,}1112$$

Señala, en cada caso, las respuestas correctas

3. El jugador de baloncesto *A* promedia 16 puntos por partido con una desviación típica de 4 puntos, y el jugador *B*, 20 puntos con una desviación típica de 6. Si se elige una muestra aleatoria de 10 partidos:

A. La media muestral de *A* se distribuye $\bar{X}_A \sim N\left(\mu = 16;\ \sigma^2 = 0{,}16\right)$.

B. La distribución de la diferencia de las medias muestrales tiene media 4 puntos y varianza 5,2.

C. La suma de las medias muestrales tiene media 36 puntos.

D. Lo más probable es que la media muestral del jugador *B* sea mayor de 4 puntos que la de A.

Las soluciones son B y C. La opción A no es correcta ya que $\sigma^2 = \frac{4^2}{10} = 1{,}6 \neq 0{,}16$. Tampoco es la opción D, pues no se indica numéricamente a qué se refiere con que un suceso sea “lo más probable” y, por lo tanto, no podemos concluir nada. La B es correcta ya que $\bar{Y} - \bar{X} \sim N\left(\mu_{\bar{Y}-\bar{X}} = 20 - 16 = 4;\ \sigma^2_{\bar{Y}-\bar{X}} = \frac{4^2}{10} + \frac{6^2}{10} = 5{,}2\right)$.

Y también lo es la opción C ya que $\bar{X} + \bar{Y} \sim N\left(\mu_{\bar{X}+\bar{Y}} = 16 + 20 = 36;\ \sigma^2_{\bar{X}+\bar{Y}} = \frac{4^2}{10} + \frac{6^2}{10} = 5{,}2\right)$.

4. Se elige una muestra aleatoria de una población con $\mu = 25$ y $\sigma = 5$. La distribución de la media muestral se puede aproximar por la de una distribución normal:

A. en cualquier caso.

B. si la distribución de la variable es normal.

C. si el tamaño muestral es grande.

D. si no se conoce la distribución exacta.

Las soluciones son B y C. Para poder aproximar la distribución de la media muestral mediante una distribución normal, debe darse al menos una de las siguientes opciones: la variable tiene que tener distribución normal o si no la tiene, el tamaño de la muestra debe ser suficientemente grande. Por tanto, las únicas respuestas que se pueden aceptar como correctas son B y C, descartando así A y D.

Elige la relación correcta entre las tres afirmaciones dadas

5. Sea *X* una variable aleatoria.

a) “*X* no sigue una distribución normal”.

b) “Se toma una muestra de tamaño 15”.

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

A. a y $b \Rightarrow c$

B. No $a \Rightarrow c$

C. $c \Rightarrow$ No a

D. No $b \Rightarrow c$

La solución es B. En general, se considera n suficientemente grande si $n \geq 30$. Por tanto, la opción b) no puede implicar nada con lo que descartamos A. Además, que el tamaño de la muestra sea distinto de 15 no implica que el tamaño vaya a ser mayor, por lo que también se descarta la opción D. Si la variable tiene una distribución normal, la distribución de la media muestra es también normal; mientras que al contrario no siempre es cierto. Por tanto, la opción C no es correcta pero sí lo es la opción B.

Señala el o los datos innecesarios para contestar

6. De una población se elige una muestra aleatoria. Para estimar la distribución de la media muestral de una determinada característica, debe conocerse:

A. la moda de la variable.

B. la varianza de la población.

C. el tamaño muestral.

D. la media de la variable.

La solución es A. La moda no es, a priori, un dato relevante a la hora de determinar la distribución de la media muestral mientras que el resto de parámetros (varianza, tamaño muestral y media) son necesarios para determinarla.

7. El 88 % de los estudiantes que se presentan a las pruebas de acceso a la universidad, supera el examen. Si se eligen *n* estudiantes, para calcular la probabilidad de que más del 90 % de ellos haya superado la prueba, mediante una distribución normal, es preciso que:

A. la distribución de los porcentajes sea normal.

B. se conozca la varianza.

C. el tamaño de la muestra sea suficientemente grande.

D. que se conozca la puntuación media de los estudiantes seleccionados.

La solución es A, B y D. Este problema se resuelve mediante el estadístico $\hat{p}$, el cual es un estimador de la proporción p. Si el tamaño de n es suficientemente grande se tiene que $\hat{p} \sim N\left(\mu = p;\ \sigma^2 = \frac{p(1-p)}{n}\right)$. Como se sabe que p = 0,88, solo es preciso que el tamaño de la muestra sea suficientemente grande. Por tanto, la opción C es el dato que sí es necesario.