ESTADÍSTICA

1.- En determinada provincia hay cuatro comarcas, C1, C2, C3 y C4, con un total de 1 500 000 personas censadas. De ellas, 300 000 residen en C1, 450 000 en C2 y 550 000 en C3. Se quiere realizar un estudio sobre las costumbres alimenticias en esa provincia basado en una muestra de 3000 personas.

a) ¿Qué tipo de muestreo deberíamos realizar si queremos que en la muestra resultante haya representación de todas las comarcas?

b) ¿Qué número de personas habría que seleccionar en cada comarca, atendiendo a razones de proporcionalidad?

c) ¿Cómo seleccionarías las personas en cada comarca?

a) Deberíamos realizar un muestreo aleatorio estratificado.

b) El número de personas que residen en C4 es:

$$1\,500\,000 - (300\,000 + 450\,000 + 550\,000) = 200\,000$$

Llamamos n_1, n_2, n_3 y n_4 al número de personas que tendríamos que seleccionar en cada comarca (C1, C2, C3 y C4, respectivamente). Entonces:

$$\frac{n_1}{300\,000} = \frac{n_2}{450\,000} = \frac{n_3}{550\,000} = \frac{n_4}{200\,000} = \frac{3\,000}{1\,500\,000}$$

Por tanto, debemos elegir:

$n_1 = 600$ personas de C1

$n_2 = 900$ personas de C2

$n_3 = 1\,100$ personas de C3

$n_4 = 400$ personas de C4

c) Dentro de cada comarca, podríamos seleccionarlos mediante un muestreo aleatorio simple, o mediante un muestreo sistemático.

2.- Halla las siguientes probabilidades en una distribución *N* (0, 1):

a) $P[z > 2{,}8]$ **b)** $P[z \le -1{,}8]$ **c)** $P[z > -1{,}8]$

d) $P[1{,}62 \le z < 2{,}3]$ **e)** $P[1 \le z \le 2]$ **f)** $P[-0{,}61 \le z \le 1{,}4]$

g) $P[-1 \le z \le 2]$ **h)** $P[-2{,}3 < z < -1{,}7]$ **i)** $P[-2 \le z \le -1]$

a) $P[z > 2{,}8] = 1 - P[z \le 2{,}8] = 1 - 0{,}9974 = 0{,}0026$

b) $P[z \le -1{,}8] = P[z \ge 1{,}8] = 1 - P[z < 1{,}8] = 1 - 0{,}9641 = 0{,}0359$

c) $P[z > -1{,}8] = P[z < 1{,}8] = 0{,}9641$

d) $P[1{,}62 \le z < 2{,}3] = P[z < 2{,}3] - P[z \le 1{,}62] = 0{,}9893 - 0{,}9474 = 0{,}0419$

e) $P[1 \le z \le 2] = P[z \le 2] - P[z \le 1] = 0{,}9772 - 0{,}8413 = 0{,}1359$

f) $P[-0{,}61 \le z \le 1{,}4] = P[z \le 1{,}4] - P[z \le -0{,}61] = P[z \le 1{,}4] - P[z \ge 0{,}61] =$

$= P[z \le 1{,}4] - (1 - P[z \le 0{,}61]) = 0{,}9192 - (1 - 0{,}7291) = 0{,}6483$.

g) $P[-1 \le z \le 2] = P[z \le 2] - P[z \le -1] = P[z \le 2] - P[z \ge 1] =$

$= P[z \le 2] - (1 - P[z \le 1]) = 0{,}9772 - (1 - 0{,}8413) = 0{,}8185$

h) $P[-2{,}3 < z < -1{,}7] = P[1{,}7 < z < 2{,}3] = P[z < 2{,}3] - P[z < 1{,}7] =$

$= 0{,}9893 - 0{,}9554 = 0{,}0339$

i) $P[-2 \le z \le -1] = P[1 \le z \le 2] = P[z \le 2] - P[z \le 1] = 0{,}9772 - 0{,}8413 = 0{,}1359$

3.- Calcula el valor de *k* (exacta o aproximadamente) en cada uno de los siguientes casos:

a) $P[z \le k] = 0{,}5$ **b)** $P[z \le k] = 0{,}8729$

c) $P[z \le k] = 0{,}9$ **d)** $P[z \le k] = 0{,}33$

e) $P[z \le k] = 0{,}2$ **f)** $P[z > k] = 0{,}12$

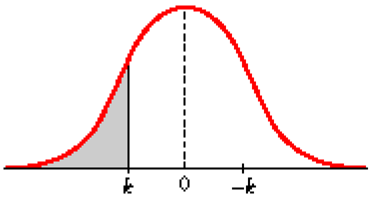
g) $P[z \ge k] = 0{,}9971$ **h)** $P[z \ge k] = 0{,}6$

a) $P[z \le k] = 0{,}5 \quad \rightarrow \quad k = 0$

b) $P[z \le k] = 0{,}8729 \quad \rightarrow \quad k = 1{,}14$

c) $P[z \le k] = 0{,}9 \quad \rightarrow \quad k = 1{,}28$

d) $P[z \le k] = 0{,}33$



$P[z \ge -k] = 0{,}33 \quad \rightarrow \quad P[z \le -k] = 1 - 0{,}33 = 0{,}67$

$\rightarrow \quad -k = 0{,}44 \quad \rightarrow \quad k = -0{,}44$

e) $P[z \le k] = 0{,}2$

$P[z \le -k] = 1 - 0{,}2 = 0{,}8 \quad \rightarrow \quad -k = 0{,}84 \quad \rightarrow \quad k = -0{,}84$

f) $P[z > k] = 0{,}12$

$P[z \le k] = 1 - 0{,}12 = 0{,}88 \quad \rightarrow \quad k = 1{,}175$

g) $P[z \ge k] = 0{,}9971$

$P[z \le -k] = 0{,}9971 \quad \rightarrow \quad -k = 2{,}76 \quad \rightarrow \quad k = -2{,}76$

h) $P[z \ge k] = 0{,}6$

$P[z \le -k] = 0{,}6 \quad \rightarrow \quad -k = 0{,}25 \quad \rightarrow \quad k = -0{,}25$

4.- En una distribución $N(18, 4)$, halla las siguientes probabilidades:

a) $P[x \le 20]$ **b)** $P[x \ge 16{,}5]$ **c)** $P[x \le 11]$

d) $P[19 \le x \le 23]$ **e)** $P[11 \le x < 25]$

a) $P[x \leq 20] = P\left[z \leq \frac{20-18}{4}\right] = P[z \leq 0,5] = 0,6915$

b) $P[x \geq 16,5] = P\left[z \geq \frac{16,5-18}{4}\right] = P[z \geq -0,38] = P[z \leq 0,38] = 0,6480$

c) $P[x \leq 11] = P\left[z \leq \frac{11-18}{4}\right] = P[z \leq -1,75] = P[z \geq 1,75] = 1 - P[z \leq 1,75] =$

$= 1 - 0,9599 = 0,0401$

d) $P[19 \leq x \leq 23] = P\left[\frac{19-18}{4} \leq z \leq \frac{23-18}{4}\right] = P[0,25 \leq z \leq 1,25] =$

$= P[z \leq 1,25] - P[z \leq 0,25] = 0,8944 - 0,5987 = 0,2957$

e) $P[11 \leq x < 25] = P\left[\frac{11-18}{4} \leq z \leq \frac{25-18}{4}\right] = P[-1,75 \leq z \leq 1,75] =$

$= P[z \leq 1,75] - P[z \leq -1,75] = P[z \leq 1,75] - P[z \geq 1,75] =$

$= 2P[z \leq 1,75] - 1 = 2 \cdot 0,9599 - 1 = 0,9198$

5.- En una distribución *N* (20, 6), tomamos muestras de tamaño 64.
a) ¿Cuál es la distribución de las medias de las muestras?
b) ¿Cuál es la probabilidad de extraer una muestra cuya media esté comprendida entre 19 y 21?

a) Las medias muestrales, $\bar{x}$, se distribuyen según una normal de media $\mu = 20$ y de desviación típica $\frac{\sigma}{\sqrt{n}} = \frac{6}{\sqrt{64}} = \frac{6}{8} = 0,75$; es decir:

$\bar{x}$ es $N(20; 0,75)$

b) $P[19 < \bar{x} < 21] = P\left[\frac{19-20}{0,75} < z < \frac{21-20}{0,75}\right] = P[-1,33 < z < 1,33] =$

$= P[z < 1,33] - P[z < -1,33] = P[z < 1,33] - (1 - P[z < 1,33]) =$

$= 2P[z < 1,33] - 1 = 2 \cdot 0,9082 - 1 = 0,8164$

6.- Se sabe que el cociente intelectual de los alumnos de una universidad se distribuye según una ley normal de media 100 y varianza 729.
a) Halla la probabilidad de que una muestra de 81 alumnos tenga un cociente intelectual medio inferior a 109.
b) Halla la probabilidad de que una muestra de 36 alumnos tenga un cociente intelectual medio superior a 109.

El cociente intelectual sigue una distruibución normal de media $\mu = 100$ y de desviación típica $\sigma = \sqrt{729} = 27$; es decir, x es $N(100, 27)$.

a) Las medias en muestras de 81 alumnos se distribuirán según una normal de media $\mu = 100$ y de desviación típica $\frac{\sigma}{\sqrt{n}} = \frac{27}{\sqrt{81}} = \frac{27}{9} = 3$; es decir, $\bar{x}$ es $N(100, 3)$. Así:

$P[\bar{x} < 109] = P\left[z < \frac{109-100}{3}\right] = P[z < 3] = 0,9987$

b) Las medias en muestras de 36 alumnos se distribuyen según una normal de media $\mu = 100$ y de desviación típica $\frac{\sigma}{\sqrt{n}} = \frac{27}{\sqrt{36}} = \frac{27}{6} = 4,5$; es decir, $\bar{x}$ es $N(100; 4,5)$. Así:

$P[\bar{x} > 109] = P\left[z > \frac{109-100}{4,5}\right] = P[z > 2] = 1 - P[z \leq 2] = 1 - 0,9772 = 0,0228$

7.- El tiempo de espera, en minutos, de los pacientes en un servicio de urgencias, es $N(14, 4)$.
a) ¿Cómo se distribuye el tiempo medio de espera de 16 pacientes?
b) En una media jornada se ha atendido a 16 pacientes. ¿Cuál es la probabilidad de que el tiempo medio de su espera esté comprendido entre 10 y 15 minutos?

a) El tiempo medio de espera, $\bar{x}$, de 16 pacientes se distribuye según una normal de media $\mu = 14$ y de desviación típica $\frac{\sigma}{\sqrt{n}} = \frac{4}{\sqrt{16}} = \frac{4}{4} = 1$; es decir $\bar{x}$ es $N(14, 1)$.

b) $P[10 < \bar{x} < 15] = P\left[\frac{10-14}{1} < z < \frac{15-14}{1}\right] = P[-4 < z < 1] =$

$= P[z < 1] - P[z < -4] = 0{,}8413 - 0 = 0{,}8413$

8.- Se sabe que el peso en kilogramos de los alumnos de Bachillerato de Madrid es una variable aleatoria, *x*, que sigue una distribución normal de desviación típica igual a 5 kg. En el caso de considerar muestras de 25 alumnos, ¿qué distribución tiene la variable aleatoria media muestral?

La variable aleatoria media muestral, $\bar{x}$, sigue una distribución normal con la misma media que la población, llamémosla μ, y con desviación típica $\frac{\sigma}{\sqrt{n}} = \frac{5}{\sqrt{25}} = \frac{5}{5} = 1$; es decir, $\bar{x}$ es $N(\mu, 1)$.

9.- En una ciudad, la altura media de sus habitantes tiene una desviación típica de 8 cm. Si la altura media de dichos habitantes fuera de 175 cm, ¿cuál sería la probabilidad de que la altura media de una muestra de 100 individuos tomada al azar fuera superior a 176 cm?

La altura en la población, x, sigue una distribución normal $N(175, 8)$. Si consideramos muestras de tamaño $n = 100$, las medias muestrales se distribuyen según una normal de media $\mu = 175$ y de desviación típica $\frac{\sigma}{\sqrt{n}} = \frac{8}{\sqrt{100}} = \frac{8}{10} = 0{,}8$; es decir, $\bar{x}$ es $N(175; 0{,}8)$. Así:

$$P[\bar{x} > 176] = P\left[z > \frac{176-175}{0{,}8}\right] = P[z > 1{,}25] = 1 - P[z \le 1{,}25] =$$

$$= 1 - 0{,}8944 = 0{,}1056$$

10.- La estatura de los jóvenes de una ciudad sigue una distribución $N(\mu, \sigma)$. Si el 90% de las medias de las muestras de 81 jóvenes están en (173,4; 175,8), halla μ□y σ□

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

El intervalo característico para las medias de las muestras de 81 jóvenes (para el 90%) es:

$$\left(\mu - 1{,}645 \cdot \frac{\sigma}{\sqrt{n}},\ \mu + 1{,}645 \cdot \frac{\sigma}{\sqrt{n}}\right)$$

El centro del intervalo es μ:

$$\mu = \frac{173{,}4 + 175{,}8}{2} = 174{,}6 = \mu$$

La semiamplitud del intervalo es:

$$1{,}645 \cdot \frac{\sigma}{\sqrt{81}} = \frac{175{,}8 - 173{,}4}{2}$$

$$1{,}645 \cdot \frac{\sigma}{9} = 1{,}2 \quad \rightarrow \quad \sigma = \frac{1{,}2 \cdot 9}{1{,}645} = 6{,}57$$

11.- Si la distribución de la media de las alturas en muestras de tamaño 49 de los niños de 10 años tiene como media 135 cm y como desviación típica 1,2 cm, ¿cuánto valen la media y la varianza de la altura de los niños de esa ciudad?

Si la media en la población es μ y la desviación típica es σ, entonces, la distribución de las medias muestrales es $N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$. Así, tenemos que:

$$\mu = 135 \text{ cm}$$

$$\frac{\sigma}{\sqrt{n}} = \frac{\sigma}{\sqrt{49}} = \frac{\sigma}{7} = 1{,}2 \text{ cm} \quad \rightarrow \quad \sigma = 1{,}2 \cdot 7 = 8{,}4 \text{ cm}$$

Por tanto, la media es $\mu = 135$ cm y la varianza es $\sigma^2 = 8{,}4^2 = 70{,}56$.

12.- Los paquetes recibidos en un almacén tienen un peso medio de 300 kg y una desviación típica de 50 kg. ¿Cuál es la probabilidad de que 25 de los paquetes, elegidos al azar, excedan el límite de carga del montacargas donde se van a meter, que es de 8 200 kg?

Sabemos que la suma de los pesos de n de esas bolsas tomadas al azar sigue una distribución normal de media $n\mu$ y de desviación típica $\sigma\sqrt{n}$, es decir:

$$\sum_{i=1}^{n} x_i \text{ es } N(n\mu, \sigma\sqrt{n})$$

En este caso:

$$\sum_{i=1}^{25} x_i \text{ es } N(25 \cdot 300;\ 50\sqrt{25}); \text{ es decir } N(7\,500;\ 250)$$

Tenemos que calcular:

$$P\left[\sum_{i=1}^{25} x_i > 8\,200\right] = P\left[z > \frac{8\,200 - 7\,500}{250}\right] = P[z > 2{,}8] =$$

$$= 1 - P[z \le 2{,}8] = 1 - 0{,}9974 = 0{,}0026$$

13.- Se ha tomado una muestra aleatoria de 100 individuos a los que se ha medido el nivel de glucosa en sangre, obteniéndose una media muestral de 110 mg/cm3. Se sabe que la desviación típica de la población es de 20 mg/cm3.

a) Obtén un intervalo de confianza, al 90%, para el nivel de glucosa en sangre en la población.

b) ¿Qué error máximo se comete con la estimación anterior?

a) Para el 90% $\rightarrow$ $1 - \alpha = 0{,}90$ $\rightarrow$ $z_{\alpha/2} = 1{,}645$

El intervalo de confianza para μ al 90% es:

$$\left(110 - 1{,}645 \cdot \frac{20}{\sqrt{100}};\ 110 + 1{,}645 \cdot \frac{20}{\sqrt{100}}\right); \text{ es decir: } (106{,}71;\ 113{,}29)$$

b) El error máximo es la mitad de la longitud del intervalo de confianza, es decir:

$$E = 1{,}645 \cdot \frac{20}{\sqrt{100}} = 1{,}645 \cdot 2 = 3{,}29$$

14.- Las medidas de los diámetros de una muestra al azar de 200 cojinetes de bolas, hechos por una determinada máquina dieron una media de 2 cm y una desviación típica de 0,1 cm. Halla los intervalos de confianza del 68,26% 95,44% y 99,73%
para el diámetro medio de todos los cojinetes.

a) Para el 68,26% $\rightarrow$ $1 - \alpha = 0{,}6826$ $\rightarrow$ $z_{\alpha/2} = 1$

El intervalo de confianza para μ es:

$$\left(2 - 1 \cdot \frac{0,1}{\sqrt{200}};\ 2 + 1 \cdot \frac{0,1}{\sqrt{200}}\right); \text{ es decir: } (1,993;\ 2,007)$$

b) Para el 95,44% $\rightarrow$ $1 - \alpha = 0,9544$ $\rightarrow$ $z_{\alpha/2} = 2$

El intervalo de confianza es:

$$\left(2 - 2 \cdot \frac{0,1}{\sqrt{200}};\ 2 + 2 \cdot \frac{0,1}{\sqrt{200}}\right); \text{ es decir: } (1,986;\ 2,014)$$

c) Para el 99,73% $\rightarrow$ $1 - \alpha = 0,9973$ $\rightarrow$ $z_{\alpha/2} = 3$

El intervalo de confianza es:

$$\left(2 - 3 \cdot \frac{0,1}{\sqrt{200}};\ 2 + 3 \cdot \frac{0,1}{\sqrt{200}}\right); \text{ es decir: } (1,979;\ 2,021)$$

15.- Sabemos que al lanzar al suelo 100 chinchetas, en el 95% de los casos, la proporción de ellas que quedan con la punta hacia arriba está en el intervalo (0,1216; 0,2784). Calcula la probabilidad *p* de que una de esas chinchetas caiga con la punta hacia arriba y comprueba que la amplitud del intervalo dado es correcta.

- p es el centro del intervalo, es decir:

$$p = \frac{0,2784 + 0,1216}{2} = 0,2 = p$$

- Veamos que la amplitud del intervalo dado es correcta:

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

El intervalo característico es:

$$\left(p - z_{\alpha/2} \cdot \sqrt{\frac{pq}{n}},\ p + z_{\alpha/2} \cdot \sqrt{\frac{pq}{n}}\right)$$

En este caso $(p = 0,2;\ q = 0,8;\ n = 100;\ z_{\alpha/2} = 1,96)$, queda:

$$\left(0,2 - 1,96 \cdot \sqrt{\frac{0,2 \cdot 0,8}{100}};\ 0,2 + 1,96 \cdot \sqrt{\frac{0,2 \cdot 0,8}{100}}\right); \text{ es decir:}$$

(0,1216; 0,2784), como queríamos probar.

16.- De 120 alumnos, la proporción de que tengan dos o más hermanos es de 48/120. Indica los parámetros de la distribución a la que se ajustarían las muestras de tamaño 30.

En muestras de tamaño $n = 30$, la proporción muestral, pr, seguiría una distribución normal de media:

$$\mu = np = 30 \cdot \frac{48}{120} = 30 \cdot 0,4 = 12$$

y de desviación típica:

$$\sigma = \sqrt{\frac{pq}{n}} = \sqrt{\frac{0,4 \cdot 0,6}{30}} = 0,089$$

Es decir, pr es $N(12;\ 0,089)$.

17.- ¿De qué tamaño conviene tomar la muestra de una línea de producción para tener una confianza del 95% de que la proporción estimada no difiere de la verdadera en más de un 4%? Se sabe, por estudios previos, que la proporción de objetos defectuosos es del orden del 0,05.

Para el 95% de confianza, $1-\alpha = 0{,}95 \rightarrow z_{\alpha/2} = 1{,}96$

El error máximo admisible es:

$$E = z_{\alpha/2} \cdot \sqrt{\frac{pr(1-pr)}{n}}$$. Buscamos n para que $E \leq 0{,}04$ (no más de un 4%):

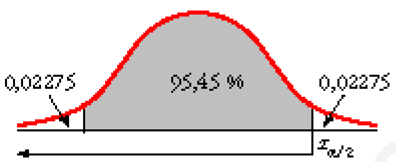
$$1{,}96 \cdot \sqrt{\frac{0{,}05(1-0{,}05)}{n}} \leq 0{,}04 \rightarrow n \geq 114{,}05$$

El tamaño mínimo de la muestra ha de ser $n = 115$.

18.- En una encuesta realizada a 800 personas elegidas al azar del censo electoral, 240 declaran su intención de votar al partido *A*.12
a) Estima, con un nivel de confianza del 95,45%, entre qué valores se encuentra la intención de voto al susodicho partido en todo el censo.
b) Discute, razonadamente, el efecto que tendría sobre el intervalo de confianza
el aumento, o la disminución, del nivel de confianza

La proporción muestral es $pr = \frac{240}{800} = 0{,}3 \rightarrow 1-pr = 0{,}7$

a) Para un nivel de confianza del 95,45%, hallamos $z_{\alpha/2}$:



$$0{,}02275 + 0{,}9545 = 0{,}97725 \rightarrow P[z \leq z_{\alpha/2}] = 0{,}97725 \rightarrow z_{\alpha/2} = z_{\alpha/2} = 2$$

El intervalo de confianza para estimar la proporción en la población es:

$$\left(pr - z_{\alpha/2} \cdot \sqrt{\frac{pr(1-pr)}{n}};\ pr + z_{\alpha/2} \cdot \sqrt{\frac{pr(1-pr)}{n}}\right)$$; en este caso queda:

$$\left(0{,}3 - 2 \cdot \sqrt{\frac{0{,}3 \cdot 0{,}7}{800}};\ 0{,}3 + 2 \cdot \sqrt{\frac{0{,}3 \cdot 0{,}7}{800}}\right)$$; es decir: (0,2676; 0,3324)

La proporción de votantes del partido *A* en la población se encuentra, con un nivel de confianza del 95,45%, entre el 26,76% y el 33,24%.

b) Si aumenta el nivel de confianza, mayor es la amplitud del intervalo; es decir, cuanto más seguros queramos estar de nuestra estimación, mayor será el error máximo admisible.

Si disminuye el nivel de confianza, disminuye la amplitud del intervalo.

19.- Una reciente encuesta, realizada en un cierto país sobre una muestra aleatoria de 800 personas, arroja el dato de que 300 de ellas son analfabetas. Para estimar la proporción de analfabetos del país hemos obtenido el siguiente intervalo de confianza: (0,3414; 0,4086)
¿Cuál es el nivel de confianza con el que se ha hecho la estimación?

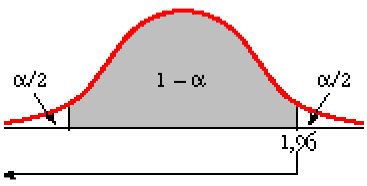
La proporción muestral es $pr = \frac{300}{800} = \frac{3}{8} \rightarrow 1 - pr = \frac{5}{8}$

El error máximo admisible es la semiamplitud del intervalo de confianza; es decir:

$$E = \frac{0{,}4086 - 0{,}3414}{2} = 0{,}0336$$

Por tanto:

$$E = z_{\alpha/2} \cdot \sqrt{\frac{pr\,(1 - pr)}{n}} \rightarrow 0{,}0336 = z_{\alpha/2} \cdot \sqrt{\frac{(3/8) \cdot (5/8)}{800}} \rightarrow z_{\alpha/2} = 1{,}96$$



$P[z \leq 1{,}96] = 0{,}9750$

$\frac{\alpha}{2} = P[z > 1{,}96] = 1 - 0{,}9750 = 0{,}025$

$\alpha = 0{,}025 \cdot 2 = 0{,}05 \rightarrow 1 - \alpha = 0{,}95$

El nivel de confianza es del 95%.