❶ Razona cuáles de los siguientes estudios estadísticos pueden efectuarse sobre toda la población y con cuáles es preferible hacer una muestra:

a) Preferencias musicales de los alumnos de tu centro escolar.
b) Preferencias deportivas de los españoles con edades comprendidas entre 16 y 19 años.
c) Número de veces, en un año, que asisten a algún espectáculo los vecinos de tu escalera.
d) Ingresos medios de los trabajadores de tu comunidad autónoma.

---oo0oo---

La necesidad de obtener una muestra suele provenir de :

- Razones económicas, es caro entrevistar a toda la población. O de falta de personal.
- Imposibilidad de realizarla a la población (población infinita o inaccesible).
- Razones de premura de tiempo que aconsejan una entrevista urgente a una muestra.
- Razones de homogeneidad de la población que aconsejan realizar un muestreo, será representativo y más barato y rápido.

a) Depende de los medios de que se disponga y el tamaño del centro se puede hacer el estudio grupo a grupo o elegir una muestra representativa de cada grupo.

b) Lo aconsejable es tomar una muestra debido al excesivo tamaño de la población.

c) A todos los vecinos, población, pues no es muy costoso ni largo.

d) Población excesivamente grande, luego es aconsejable tomar una muestra.

❷ Explica por qué la muestra elegida en cada uno de los casos siguientes no es representativa:

a) Para conocer el porcentaje de españoles a los que les gusta el fútbol, un periódico deportivo realiza una encuesta entre sus lectores.
b) Para conocer los hábitos de lectura de los españoles, una empresa de encuestas entrevista a 100 universitarios.
c) Para conocer los gustos musicales de los españoles, se encuesta a 200 personas a la salida de un concierto.

---oo0oo---

a) No será representativa pues al realizarla entre los lectores de periódicos deportivos estará sesgada, ya les interesa el deporte a los lectores de prensa deportiva

b) Tampoco es representativa pues, normalmente, un universitario tendrá más afición a la lectura que el resto.

c) No representativa pues, si han ido al concierto, una proporción elevada de personas, es por que les gusta la actuación y los gustos musicales irán en ese sentido.

❸ En un centro escolar estudian 350 alumnos de ESO y 150 alumnos de Bachillerato. Explica cómo se puede seleccionar una muestra de 50 alumnos usando los tres métodos de muestreo aleatorio descritos anteriormente.

---oo0oo---

Muestreo aleatorio sistemático :

Asignamos un número a las 500 alumnos, por orden alfabético por ejemplo, y después por cualquier procedimiento aleatorio (una tabla de números aleatorios, la calculadora o el ordenador) seleccionamos 50.

Muestreo aleatorio sistemático:

- Tamaño de la población N = 500.
- Tamaño de la muestra n = 50.
- Tamaño del incremento k = N/n = 10.

1) Una vez numerados del 1 al 500, elegimos un número al azar entre los diez primeros ($1 \leq n_1 \leq 10$)

2) Los otros 40 individuos de la población serán los números :

$n_1 + 10,\ n_1 + 20,\ n_1 + 30,\ \dots,\ n_1+480,\ n_1+490$

Muestreo aleatorio estratificado

Los estratos van a ser el nivel de estudios, es decir hacemos dos estratos los 350 que estudian ESO y los 150 que estudian Bachillerato.

1) *Afijación homogénea*: como hay dos estratos seleccionamos 25 individuos de cada estrato mediante muestreo aleatorio simple o por el sistemático.

2) *Afijación proporcional*: como en el primer estrato hay el 70% (350/500 = 0'7) de los individuos y, en el segundo, el 30 % seleccionamos 35 individuos (50 · 0'7) y 15 (50 · 0'3) del segundo, mediante algunos de los procedimientos anteriores.

3) *Afijación óptima* : Si son conocidos la desviación típica y el coste, se hace un reparto entre los dos estratos proporcional al tamaño y a la desviación típica e inversamente proporcional al coste.

❁ ❁ ❁ ❁ ❁❁ ❁ ❁ ❁ ❁ ❁

❹ Una fábrica produce ciertas piezas con una longitud media de 10 cm y una desviación típica de 1 cm.

a) ¿Cuál es la probabilidad de que la longitud media en una muestra de 50 piezas sea superior a 10,5 cm?

b) Si se toman 25 muestras de 50 piezas cada una, ¿en cuántas cabe esperar que la longitud media esté comprendida entre 9,8 cm y 10,3 cm?

---oo0oo---

$\mu = 10$ cm , $\sigma = 1$ cm , $n = 50$

Como el tamaño n = 50 > 30 podemos tomar la distribución muestral como una normal de media $\mu(x) = \mu = 10$ cm y desviación típica :

$$\sigma(x) = \frac{\sigma}{\sqrt{n}} = \frac{1}{\sqrt{50}} = 0'1414$$

Es decir N(10, 0'1414)

Luego:

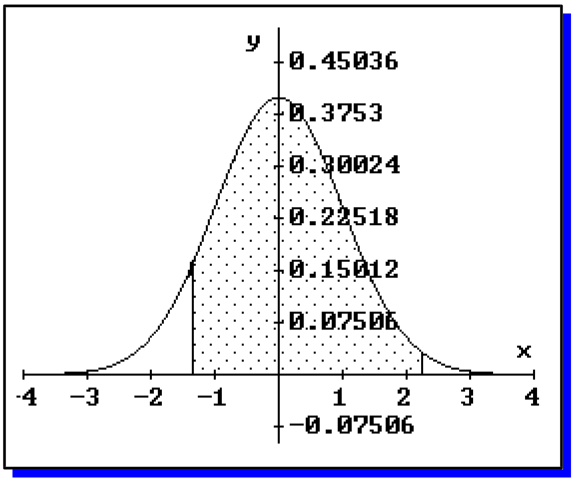
$$p(\bar{x} > 10'5) = p\left(z > \frac{10'5-10}{0'1414}\right) = p(z > 3'54) = 1 -$$

$p(z \leq 0'354) = 1- \Phi(3'54) = 1 - 0'9998 =$ **0'0002**.

b) Hemos de hallar ahora:

$$p(9'8 \leq \bar{x} \leq 10'3) = p\left(\frac{9'8-10}{0'1414} \leq z \leq \frac{10'3-10}{0'1414}\right) =$$

$p(-1'41 \leq z \leq 2'12) = p(z \leq 2'12) - p(z \leq -1'41) = p(z \leq 2'12) - [1 - p(z \leq -1'41)] = p(z \leq 2'12) + p(z \leq -1'41) - 1 = \Phi(2'12) + \Phi(1'41) - 1 = 0'9826 + 0'9207 - 1 = 0'9033$.



Luego nº de muestras = 25· 0'9033 = 22'58 ≈ **23 muestras tienen una media comprendida entre 9'8 cm y 10'3 cm.**

❁ ❁ ❁ ❁ ❁❁ ❁ ❁ ❁ ❁ ❁

❺ Al 75 % de los jóvenes de una ciudad les gusta el cine. Si seleccionamos 25 muestras de 100 jóvenes cada una, ¿en cuántas cabe esperar que el porcentaje de jóvenes cinéfilos esté comprendido entre el 70 % y el 80 %? ¿Y si las muestras fuesen de 1 000 jóvenes?

- Compara ambos resultados y extrae conclusiones.

---oo0oo---

Población p = 0'75.

Estamos en una distribución muestral de proporciones :

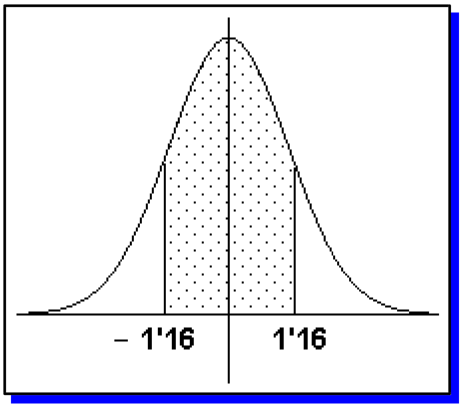
1) n = 100, luego la distribución muestral será una N (μ(p), σ(p)), en donde:

$$\mu(p) = p = 0'75$$
$$\sigma(p) = \sqrt{\frac{p\cdot(1-p)}{n}} = \sqrt{\frac{0'75\cdot 0'25}{100}} = 0'043$$

Luego :

$$p(0'7 \le \hat{p} \le 0'8) = p\left(\frac{0'7-0'75}{0'043} \le z \le \frac{0'8-0'75}{0'043}\right) =$$

p(-1'16 ≤ z ≤ 1'16) = p(z ≤ 1'16) - p(z ≤-1'16) = p(z ≤ 1'16) - [1 - p(z ≤1'16)] = 2 p(z ≤ 1'16) - 1 = 2· 0'8770 - 1 = 0'754



Entonces las muestras serán 25 · 0'754 = 18'85 ≈ **19 muestras están comprendidas entre el 70 y el 80 %.**

2) n = 1 000, luego la distribución muestral será una N (μ(p), σ(p)), en donde:

$$\mu(p) = p = 0'75$$
$$\sigma(p) = \sqrt{\frac{p\cdot(1-p)}{n}} = \sqrt{\frac{0'75\cdot 0'25}{1000}} = 0'013$$

Luego :

$$p(0'7 \le \hat{p} \le 0'8) = p\left(\frac{0'7-0'75}{0'0137} \le z \le \frac{0'8-0'75}{0'0137}\right) =$$

p(-3'65 ≤ z ≤ 3'65) = p(z ≤ 3'65) - p(z ≤ -3'65) = p(z ≤ 3'65) - [1 - p(z ≤ 3'65)] = 2 p(z ≤ 3'65) - 1 = 2· 0'9999 - 1 = 0'998

Luego habrá 25 · 0'998 = 24'995 ≈ **25 muestras** (¡ todas las muestras).

Se observa que al aumentar el tamaño de la muestra el número de muestras comprendido en el intervalo del 70 al 80 % aumenta pues de cada vez el estadístico muestral se acerca más al parámetro poblacional .

❁ ❁ ❁ ❁ ❁❁ ❁ ❁ ❁ ❁ ❁

❻ En el último año, el peso de los recién nacidos de dos maternidades A y B se ha distribuido normalmente (en la A, con media 3400 g y desviación típica 100 g; en la B, con media 3200 g y desviación típica 250 g). Halla la probabilidad de que, tomando muestras de 10 recién nacidos de cada maternidad, el peso medio de A supere al de B en más de 150 g.

---oo0oo---

Poblaciones :N(3400, 100)$_A$, N(3200, 250)$_B$

Muestras : $n_A = n_B = 10$

Distribución muestral

Tipo = de diferencia de medias

Media= $\mu(D) = \mu_A - \mu_B$ = 3400 - 3200 = 200

$$DT=\sigma(D)=\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}} = \sqrt{\frac{100^2}{10}+\frac{250^2}{10}} = 85'147$$

Luego es una N(200, 85'147)

Se ha de hallar :

$$p(\bar{X}_A - \bar{X}_B > 150) = p\left(z > \frac{150-200}{85'147}\right) = p(z > -0'59)$$

= p(z ≤ 0'59) = Φ(0'59) = **0'7224 (72'24%)** .

❁ ❁ ❁ ❁ ❁❁ ❁ ❁ ❁ ❁ ❁

❼ Considera la población formada por los números 1, 2, 3 y 4.

a) Calcula el parámetro poblacional μ.
b) Selecciona por muestreo aleatorio simple con reemplaza miento una muestra de tamaño 3 y, a partir de ella, estima μ, usando la media y usando Me. Compara tus estimaciones con el valor real y con las estimaciones obtenidas por tus compañeros.
c) Considera todas las posibles muestras de tamaño 3 seleccionadas como se indica en **b)**, obtén los dos estimadores antes utilizados y halla su distribución de muestreo. Comprueba que los dos estimadores son insesgados y que el primero es más eficiente que el segundo.

---oo0oo---

a) $\mu = \frac{\sum_{i=1}^{i=4} x_i}{N} = \frac{1+2+3+4}{4} = \frac{10}{4} = 2'5$

b) Vamos a seleccionar 3 individuos de la población mediante un m.a.s. generando con la calculadora números aleatorios y tomando los 3 primeros que aparezcan:
0'**2**24 0'087 0'**3**02

Ahora hallamos la media y la mediana de la muestra formada por 2,3 y 4 :

$$\overline{X} = \frac{2+3+4}{3} = \frac{9}{3} = 3$$

Me = 3 (el término central de la media ordenada)
Vemos que son valores próximos pero diferentes del real (2'5).

c) Vamos a obtener las posibles muestras de tamaño 3 con reemplazamiento (que son $VR_{4,3} = 4^3 = 64$) y las tabulamos con su media y mediana. Después hallaremos la media y la Me de las 64 muestras . La esperanza para ver su sesgo y su varianza para estudiar la eficiencia de cada una de las medidas :

Muestra	$\overline{X}$	Me	Muestra	$\overline{X}$	Me
1 1 1	1	1	1 3 3	2'33	3
1 1 2	1'33		3 1 3		
1 2 1			3 3 1		
2 1 1			1 3 4	2'66	
1 1 3	1'66		1 4 3		
1 3 1			3 1 4		
3 1 1			4 1 3		
1 1 4	2		4 3 1		
1 4 1			3 4 1		
4 1 1			2 3 3	2'66	
1 2 2	1'66	2	3 2 3		
2 1 2			3 3 2		
2 2 1			2 3 4	3	
1 2 3	2		2 4 3		
1 3 2			3 2 4		
2 1 3			3 4 2		
2 3 1			4 2 3		
3 1 2			4 3 2		
3 2 1			3 3 3	3	
1 2 4	2'33		3 3 4	3'33	
1 4 2			3 4 3		
2 1 4			4 3 3		
2 4 1			1 4 4	3	4
4 1 2			4 1 4		
4 2 1			4 4 1		
2 2 2	2		2 4 4	3'33	
2 2 3	2'33		4 2 4		
2 3 2			4 4 2		
3 2 2			3 4 4	3'66	
2 2 4	2'66		4 3 4		
2 4 2			4 4 3		
4 2 2			4 4 4	4	

Para hallar la media y desviación típica de las 64 muestras hacemos otra tabla :

$\overline{X}_i$	f_i	$x_i \cdot f_i$	x_i^2	$x_i^2 \cdot f_i$
1	1	1	1	1
1'33	3	4	16/9	16/3
1'66	6	10	25/9	50/3
2	10	20	4	40
2'33	12	28	49/9	196/3
2'66	12	32	64/9	256/3
3	10	30	9	90
3'33	6	20	100/9	200/3
3'66	3	11	121/9	121/3
4	1	4	16	16
Suma	64	160		426'6

Luego :

$$\mu(\overline{X}) = \frac{160}{64} = 2'5 = \mu$$

$$\sigma(\overline{X}) = \sqrt{\frac{426'66}{64} - 2'5^2} = 0'6454$$

De paso podemos apreciar que esta desviación típica es :

$$\sigma(\overline{X}) = \frac{\sigma}{\sqrt{3}} = \frac{\sqrt{\frac{1+4+9+16}{4} - 2'5^2}}{\sqrt{3}} = 0'6454$$

hagamos lo mismo para la mediana :

$Me(x_i)$	f_i	$x_i \cdot f_i$	x_i^2	$x_i^2 \cdot f_i$
1	10	10	1	10
2	22	44	4	88
3	22	66	9	198
4	10	40	16	160
Suma	64	160		456

Luego :

$$\mu(Me) = \frac{160}{64} = 2'5 = \mu$$

$$\sigma(Me) = \sqrt{\frac{456}{64} - 2'5^2} = 0'9354$$

Vemos que los dos son insesgados pues sus medias son iguales a la de la población, además la media es más eficiente que la mediana por tener menor desviación típica.

❋ ❋ ❋ ❋ ❋❋ ❋ ❋ ❋ ❋ ❋

❽ Las puntuaciones obtenidas en ciertas oposiciones siguen una distribución N (μ, 25). Si en una muestra de 60 candidatos la media fue 85 puntos, halla intervalos de confianza para la puntuación media con niveles de confianza del 99 %, del 95 % y del 90 %. Fíjate en la amplitud de los intervalos obtenidos y extrae conclusiones.

---oo0oo---

Población : N(μ, 25)

Muestra : n = 60, $\overline{X} = 85$

❀ Distribución muestral :

Tipo = de medias, normal, ya que n =60 > 30.

Media = $\mu(\overline{X}) = \mu$, desconocida

D.T. = $\sigma(\overline{X}) = \frac{\sigma}{\sqrt{n}} = \frac{25}{\sqrt{60}} = 3'2275$

$$\text{Intervalo} \equiv (\overline{X} \pm z_{\frac{\alpha}{2}} \sigma(\overline{X}))$$

Necesitamos los valores de z y para hallarlo se nos da tres niveles de confianza diferentes :

1) nivel de confianza = 99 % ⇒1 − α = 0’99 ⇒ α = 0’01⇒ α/2 = 0’005 , buscamos en la tabla normal que z deja por encima de si un 0’005, es decir el valor de z que deja por debajo un 0’995, tomamos $z_{0'995}$ = 2’58.

ya podemos hallar el Intervalo de confianza para la media poblacional :

$$I.C \equiv (\overline{X} \pm z_{\frac{\alpha}{2}} \sigma(\overline{X})) = (85 \mp 2'58 \cdot 3'2275) =$$
= (85± 8’32695) = **[76’67, 93’33]**

2) nivel de confianza = 95 % ⇒ 1 – α = 0’95 ⇒ α = 0’05⇒ α/2 = 0’025 , buscamos en la tabla normal que z deja por encima de si un 0’025, es decir el valor de z que deja por debajo un 0’975, tomamos $z_{0'975}$ = 1’96.

Ya podemos hallar el Intervalo de confianza para la media poblacional :

$$I.C \equiv (\overline{X} \pm z_{\frac{\alpha}{2}} \sigma(\overline{X})) = (85 \mp 1'96 \cdot 3'2275) =$$
= (85± 6’33) = **[78’67, 91’33]**

3) nivel de confianza = 90 % ⇒1 – α = 0’9 ⇒ α = 0’1⇒ α/2 = 0’05 , buscamos en la tabla normal que z deja por encima de si un 0’05, es decir el valor de z que deja por debajo un 0’95, tomamos $z_{0'95}$ = 1’65.

Ya podemos hallar el Intervalo de confianza para la media poblacional :
$I.C \equiv (\bar{X} \pm z_{\frac{\alpha}{2}}\, \sigma(\bar{X})) = (85 \mp 1'65 \cdot 3'2275) =$
= (85±4’84) = **[80’16, 89’84]**

Observamos que al disminuir el nivel de confianza el intervalo de confianza disminuye de amplitud, lo que es lógico pues cuanta mayor confianza se desee, más amplio ha de ser el intervalo para aumentar la seguridad de que la media esté en su interior.

❁ ❁ ❁ ❁ ❁❁ ❁ ❁ ❁ ❁ ❁

❾ Para estudiar la proporción de estudiantes que practican determinado deporte, se toma una muestra de tamaño 300. El resultado obtenido es que 210 lo practican. Calcula el intervalo de confianza para la proporción p con nivel de confianza del 98 %.

---oo0oo---

- **Población** : desconocida.
- **Muestra :** n = 300, p = 210/300 = 0’7.
- **Distribución muestral :**
  - Tipo : de proporciones.
  - Media : $\mu(p) = P \approx p = 0'7$, ya que como el tamaño es grande, al no conocer la proporción poblacional (P) la aproximamos por la muestral (p)
  - Desviación típica:

$$\sigma(p) = \sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{0'7 \cdot 0'3}{300}} = 0'02646$$

- **Intervalo de confianza para la proporción poblacional :**

$$\left(p - z_{\frac{\alpha}{2}} \cdot \sigma(p) \le P \le p + z_{\frac{\alpha}{2}} \cdot \sigma(p)\right)$$

Para su cálculo necesitamos z, como el nivel de confianza pedido es del 98 % , 1 - α = 0’98 ⇒ α = 0’02 ⇒ α/2 = 0’01 , hemos de buscar en la tabla el z que deja por debajo 0’99, es decir el $z_{0'99} = 2'33$

I.C=(0’7-2’33·0’02646≤P≤0’7+2’33 · 0’02646)

I.C. = (0’7 - 0’062 ≤ P ≤ 0’7 + 0’062) = [0’64, 0’762]

❁ ❁ ❁ ❁ ❁❁ ❁ ❁ ❁ ❁ ❁

❿ Para saber si una moneda está o no trucada, se efectúan 700 lanzamientos y se obtienen 425 caras. Halla el intervalo de confianza para la proporción de caras obtenidas en sucesivos lanzamientos, con un nivel de confianza del 95 %. Según el resultado obtenido, razona si es de esperar o no que la moneda esté trucada.

---oo0oo---

- **Población** : desconocida.
- **Muestra :** n = 700, p = 425/700 = 0’61.
- **Distribución muestral :**
  - Tipo : de proporciones.
  - Media : $\mu(p) = P \approx p = 0'61$, ya que como el tamaño es grande, al no conocer la proporción poblacional (P) la aproximamos por la muestral (p)
  - Desviación típica:

$$\sigma(p) = \sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{0'61 \cdot 0'39}{700}} = 0'01844.$$

- **Intervalo de confianza para la proporción poblacional :**

$$\left(p - z_{\frac{\alpha}{2}} \cdot \sigma(p) \le P \le p + z_{\frac{\alpha}{2}} \cdot \sigma(p)\right)$$

Para su cálculo necesitamos z, como el nivel de confianza pedido es del 95 % , 1 - α = 0’95 ⇒ α = 0’05 ⇒ α/2 = 0’025 , hemos de buscar en la tabla el z que deja por debajo 0’975, es decir el $z_{0'975} = 1'96$

I.C. = (0'61 - 1'96· 0'01844 $\leq$ P $\leq$ 0'61 + 1'96 · 0'01844) = [0'57, 0'65]

Si no estuviese trucada su probabilidad sería ½ = 0'5 pero este valor no está en el intervalo de confianza hallado, luego **es de sospechar que la moneda esté trucada.**

❁ ❁ ❁ ❁ ❁❁ ❁ ❁ ❁ ❁ ❁

❶❶ Una muestra de 50 bombillas de la marca A dio una vida media de 1 500 h y una desviación típica de 100 h. Una muestra de 65 bombillas de la marca B dio una vida media de 1400 h y una desviación típica de 150 h. Halla el intervalo de confianza para la diferencia de medias de ambas marcas, con un nivel de significación del 6 %. Si ambas marcas venden las bombillas al mismo precio, ¿cuáles conviene comprar? ¿Por qué?

---oo0oo---

❖ **Poblaciones** : desconocidas.

❖ **Muestras :** $n_A = 50$, $\bar{x}_A = 1\,500$ hr , $s_A = 100$ hr, $n_B = 65$, $\bar{x}_B = 1\,400$ hr , $s_B = 150$ hr .

❖ **Distribución muestral :**

✲ Tipo : de diferencia de medias.
✲ Media : $\mu(D) = \mu_A - \mu_B$, desconocida
✲ Desviación típica:

$$\sigma(D) = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

Como no conocemos la desviaciones poblacionales usamos las muestrales al ser los tamaños mayores que 30:

$$\sigma_1^2 = \frac{n_1}{n_1-1} s_1^2 = \frac{50}{49} 100^2 = 10204'1$$

$$\sigma_2^2 = \frac{n_2}{n_2-1} s_2^2 = \frac{65}{64} 150^2 = 22851'6$$

Luego :

$$\sigma(D) = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} = \sqrt{\frac{10204'1}{50} + \frac{22851'6}{65}} = 23'57$$

❖ **Intervalo de confianza para la diferencia de medias poblacional :**

$$\left((\bar{x}_A - \bar{x}_B) \pm z_{\frac{\alpha}{2}} \cdot \sigma(D)\right)$$

Para su cálculo necesitamos z, como el nivel de significación es del 6 % $\Rightarrow \alpha = 0'06 \Rightarrow \alpha/2 = 0'03$, hemos de buscar en la tabla el z que deja por debajo 0'97, es decir el $z_{0'97} = 1'89$

I.C = [(1 500 - 1 400) ±1'89·23'57)

I.C. = [100 ± 44'55] = [55'45, 144'55]

Esto significa que cabe esperar que las bombillas de la marca A duren al menos 55'45 hr. más que las de marca B, luego **convendría comprar las de la marca A pues presumiblemente, por término medio, duran unas 55 horas más**.

❁ ❁ ❁ ❁ ❁❁ ❁ ❁ ❁ ❁ ❁

❶❷ Queremos saber el tiempo medio diario de estudio de los alumnos de Bachillerato españoles. En una muestra de 100 alumnos se obtuvo una desviación típica de 53 minutos. ¿Qué tamaño muestra¡ hay que tomar para que el error máximo sea de 5 minutos, con una significación del 5 %?

---oo0oo---

n = 100 , s = 53 min, E = 5 min, $\alpha = 0'05$

sabemos que el error cometido es como máximo :

$$E = z_{\frac{\alpha}{2}} \cdot \sigma(\bar{X}) = z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}} \Leftrightarrow n = \left(\frac{z_{\frac{\alpha}{2}} \cdot \sigma}{E}\right)^2$$

Como no sabemos la desviación típica poblacional, la estimamos a partir de la muestral :

$$\sigma^2 = \frac{n}{n-1}s^2 = \frac{100}{99}53^2 = 2837'\widehat{37}$$

Luego, como α/2 = 0'025 $z_{0'975}$ = 1'96:

$$n = \frac{1'96^2 \cdot 2837'\overline{37}}{5^2} = 436'002$$

Tomamos una muestra de al menos 436 individuos.

❁ ❁ ❁ ❁ ❁❁ ❁ ❁ ❁ ❁ ❁

❶❸ En una muestra de 250 habitantes de una zona, 183 se manifestaron favorables a la apertura de un nuevo supermercado. ¿De qué tamaño deberá tomarse la muestra para que, con un nivel de confianza del 99 %, la proporción muestra¡ y la poblacional no difieran en más de 0,02?

---oo0oo---

p = 183/250= 0'732, 1 - α = 0'99, E = 0'02

$$n = \left(\frac{z_{\frac{\alpha}{2}} \cdot \sigma}{E}\right)^2$$

Necesitamos :

❍ z, como 1 - α = 0'99 ⇨ α = 0'01 ⇨ α/2 = 0'005 ⇨ $z_{0'995}$ = 2'58

❍ La desviación típica poblacional que es desconocida la hallamos a través de la muestral :

$$\sigma = P \cdot Q \simeq p \cdot q = 0'732 \cdot 0'268 = 0'1962$$

Luego:

$$n = \left(\frac{z_{\frac{\alpha}{2}} \cdot \sigma}{E}\right)^2 = \frac{2'58^2 \cdot 0'1962}{0'02^2} = 3264'56$$

Hemos de tomar al menos 3 265 individuos en la muestra.

❁ ❁ ❁ ❁ ❁❁ ❁ ❁ ❁ ❁ ❁

Resolución de ejercicios y problemas

❶ Repite los apartados b y c del ejercicio resuelto efectuando tus propios sorteos, y elabora tus propias conclusiones.

---oo0oo---

b) k = 3 , con la calculadora y la tecla [RAN#] vamos obteniendo números aleatorios hasta que salga un número comprendido entre 1 y 50 , resulta ser el 1, y ahora seleccionamos sistemáticamente cada 3 alumnos los 15 :

Nº	x	Nº	x	Nº	x
①	6	①⑥	6	③①	4
④	9	①⑨	8	③④	5
⑦	5	②②	7	③⑦	9
⑩	8	②⑤	7	④⓪	6
①③	10	②⑧	10	④③	9

$$\bar{x} = \frac{\sum_{i=1}^{15} x_i}{15} = \frac{109}{15} = 7'2\hat{6}$$

Hay una diferencia con respecto de la población de 6'16 - 7'2666.. = -1'1 puntos que puede deberse, fundamentalmente al tamaño pequeño de la muestra.

c) Los tamaños de cada estrato se han calculado en libro proporcionalmente al tamaño de cada uno y son : n_c = 13 , n_T = 10 y n_H = 17.

Realizamos, dentro de cada estrato, un m.a.s. generando números aleatorios ;

Científico		Tecno		Humanístico	
Nº	x	Nº	x	Nº	x
9	6	28	10	45	7
10	8	28	10	48	5
12	5	19	8	33	2
12	5	21	4	34	5
12	5	25	7	48	5
4	9	26	7	31	4
7	5	20	5	43	9
8	6	25	7	40	6
12	5	18	5	41	5
13	10	22	7	43	9
3	5			46	6
10	8			46	6
12	5			32	6
				33	2
				30	5
				43	9
				31	4

Hallamos ahora la media :

$$\bar{x} = \frac{253}{40} = 6'325$$

Una resultado más próximo, seguramente debido al aumento del tamaño de la muestra , no al tipo de muestreo.

❁ ❁ ❁ ❁ ❁❁ ❁ ❁ ❁ ❁ ❁

❷ Repite los apartados b y c del ejercicio resuelto tomando en b una muestra de tamaño 45 y en c una de tamaño 20. Extrae conclusiones.

---oo0oo---

b) k = 50/45 = 1'1≈ 1 , con la calculadora y la tecla [RAN#] vamos obteniendo números aleatorios hasta que salga un número comprendido entre 1 y 50 , resulta ser el 9, y ahora seleccionamos sistemáticamente los 15 :

Nº	x	Nº	x	Nº	x
9	6	24	3	39	6
10	8	25	7	40	6
11	9	26	7	41	5
12	5	27	5	42	5
13	10	28	10	43	9
14	6	29	5	44	6
15	7	30	5	45	7
16	6	31	4	46	6
17	6	32	6	47	3
18	5	33	2	48	5
19	8	34	5	49	7
20	5	35	5	50	9
21	4	36	5	1	6
22	7	37	9	2	7
23	5	38	7	3	5

$$\bar{x} = \frac{\sum_{i=1}^{15} x_i}{15} = \frac{274}{45} = 6'09$$

Hay una diferencia con respecto de la población de 6'16 - 6'09 = 0'07 puntos, diferencia pequeña pues ya el tamaño es similar al de la población.

c) Los tamaños de cada estrato son : n_c = 16·20/50 = 6'4 ≈ 7 , n_T = 13·20/50 = 5'2 ≈ 5 y n_H = 21·20/50 = 8'4 ≈ 8.

Realizamos, dentro de cada estrato, un m.a.s. generando números aleatorios ;

Científico		Tecno		Humanístico	
Nº	x	Nº	x	Nº	x
9	6	21	4	42	5
8	6	22	7	45	7
10	8	26	7	39	6
12	5	27	5	49	7
12	5	28	10	49	7
12	5			34	5
4	9			37	9
				44	6

Hallamos ahora la media :

$$\bar{x} = \frac{129}{20} = 6'45$$

Una resultado más alejado, seguramente debido al disminuir el tamaño de la muestra , no al tipo de muestreo.

❁ ❁ ❁ ❁ ❁❁ ❁ ❁ ❁ ❁ ❁

❸ Un agricultor vende tomates con un peso medio de 75 g y una desviación típica de 15 g.

a) Halla la probabilidad de que una muestra de 100 tomates tenga un peso total inferior a 7 kg.

b) Se sabe que el 70 % de los lotes de 100 tomates que vende ese agricultor supera cierto peso total x. ¿Cuál es el valor de x?

---oo0oo---

a)

Población ➢ $\mu = 75$ g y $\sigma = 15$ g

Muestra ➢ n = 100 , $\bar{x} = \frac{7000}{100} = 70$

Distribución Muestral :

- Tipo : de medias.
- Media : $\mu(x) = \mu = 75$ g.
- Desviación típica :

$$\sigma(\bar{x}) = \frac{\sigma}{\sqrt{n}} = \frac{15}{\sqrt{100}} = \frac{15}{10} = 1'5$$

Como n = 100 > 30 Será una N(75, 1'5) la distribución muestral de medias.

$$p(\bar{x} < 70) = p(z < \tfrac{70-75}{1'5}) = p(z < -3'\hat{3}) =$$

$= 1 - p(z \leq 3'33) = 1 - \Phi(3'33) = 1 - 0'996 =$ $=$ **0'0004** (un 0'4 o/oo)

b) n = 100

Se sabe que p(z > m) = 0'7 , buscamos en la tabla a que valor de z le corresponde, teniendo en cuenta que ha de ser negativo (a cada lado de la media hay el 50 %) es decir hay que buscar el z tal p(z < m) = 0'7 y m será negativo :

0'52 = - m, pero la distribución muestral, como hemos visto es N(75, 1'5) , luego:

$$-m = -\frac{\bar{x}-75}{1'5} = 0'52 \Leftrightarrow \bar{x} - 75 = -0'52 \cdot 1'5$$

es decir $\bar{x} = 75 - 0'78 = 74'22\text{kg}$

❁ ❁ ❁ ❁ ❁❁ ❁ ❁ ❁ ❁ ❁

❹ Una empresa desea averiguar el porcentaje de cerillas defectuosas de una partida. Para ello revisa 100 de las cerillas, y encuentra 8 defectuosas.

a) Con nivel de significación del 6 %, ¿qué error máximo puede cometerse generalizando el resultado obtenido a toda la partida?

b) ¿Cuántas cerillas más se deben revisar si queremos conseguir que el error máximo cometido sea de un 3 %, con ese mismo nivel de significación?

---oo0oo---

a)
n = 100, p = 8/100 = 0'08, q = 1- p = 0'92

Como $\alpha = 0'06 \Rightarrow \alpha/2 = 0'03 \Rightarrow z_{0'97} = 1'89$

El error máximo será :

$$E = z_{\frac{\alpha}{2}} \cdot \sqrt{\frac{p \cdot q}{n}} = 1'89 \cdot \sqrt{\frac{0'08 \cdot 0'92}{100}} = 0'0512$$

b)
El error máximo ha de ser E = 0'03 (3 %)

$$E = 1'89 \cdot \sqrt{\frac{0'08 \cdot 0'92}{n}} = 0'03 \Leftrightarrow n = \frac{1'89^2 \cdot 0'0736}{0'03^2} =$$

= **292'12, luego hay que tomar una muestra de al menos 293 cerillas para que el error máximo sea del 3 %.**

❁ ❁ ❁ ❁ ❁❁ ❁ ❁ ❁ ❁ ❁