Actividades

Cuestiones

❶ Una máquina envasa bolsas de patatas fritas con un peso medio de 100 g. Seleccionamos 10 bolsas al azar y observamos un peso medio de 98 g. Define los conceptos de población, muestra, parámetro y estadístico, e identifícalos en este ejemplo.

---oo0oo---

✻ **Población** : Conjunto de elementos (individuos) finito o infinito que comparten una o más características que los definen y diferencian del resto, y que son el objeto del estudio, el universo estadístico. Todas las bolsas que salgan de la máquina.

✻ **Muestra** : Es un subconjunto de la población, debidamente seleccionado, que se somete al estudio en representación de la población. Las n = 10 bolsas que hemos seleccionado al azar.

✻ **Parámetros** : números que caracterizan e informan sobre aspectos de la población (valor central, dispersión, simetría, curtosis, etc). El peso medio de todas las bolsas m = 100g.

✻ **Estadísticos** : esos números o valores pero hallados sobre los n individuos de la muestra. El peso medio de las 10 bolsas de la muestra $\bar{x} = 98$g.

❁ ❁ ❁ ❁ ❁❁ ❁ ❁ ❁ ❁ ❁

❷ Di qué ventajas y qué inconvenientes presentan los diferentes tipos de muestreo aleatorio descritos en la unidad.

---oo0oo---

✲ **Muestreo aleatorio simple**:

◎ *Inconvenientes :* El proceso de selección, si no se usa el ordenador u otro procedimiento rápido puede ser dificultoso.

◎ *Ventajas* : Proporciona una muestra representativa si el tamaño es suficiente y permite las probabilidades de las muestras posibles y de que cada individuo sea seleccionado.

✲ **Muestreo aleatorio sistemático**:

◎ *Inconvenientes* : Si la población tiene cierta regularidad y el sistema coincide con ella los resultados obtenidos no serán representativos.

◎ *Ventajas*: El procedimiento de selección se simplifica.

✲ **Muestreo aleatorio estratificado**:

◎ *Inconvenientes*: Requiere que hay alguna cualidad diferencial para los estratos y requiere clasificación previa de los individuos de la población , según esa característica, en estratos.

◎ *Ventajas* : Cada estrato puede ser representativo del total si está bien clasificada la población y la muestra es suficiente.

❁ ❁ ❁ ❁ ❁❁ ❁ ❁ ❁ ❁

❸ Di qué es más probable y explica por qué:

a) Que la estatura de un hombre supere 1,8 m.

b) Que la estatura media de una muestra de 100 hombres supere los 1,8 m.

---oo0oo---

Depende de la media de la población, si esta tiene una estatura inferior a 1'8, es claramente más probable la a) pues cuanto más individuos tomemos más se acercarán a la media poblacional y esta es menor que 1'8.

Si, la media poblacional es superior a 1'8 sigue puede que se inviertan y sea más probable si tomamos más individuos que la media de la muestra supere los 1'8.

❁ ❁ ❁ ❁ ❁❁ ❁ ❁ ❁ ❁ ❁

❹ ¿Cuáles son las tres característi-cas que debe reunir un buen estimador?

---oo0oo---

❍ **Insesgado** : debe coincidir la esperanza del valor estimado con el valor a estimar.

❍ **Eficiente** : Que tenga poca dispersión respecto al real, lo mide la desviación típica.

❍ **Consistente** : si la probabilidad de que el estimador se aproxime al verdadero valor del parámetro poblacional tiende a 1 a medida que aumenta el tamaño de la muestra.

Es deseable que tenga una cuarta carac-terística :

❍ Suficiente : El estimador debe tomar en consideración la mayor parte posible de la información proporcionada por la muestra

❁ ❁ ❁ ❁ ❁❁ ❁ ❁ ❁ ❁ ❁

❺ Explica las ventajas y los inconve-nientes que presenta la estimación por intervalos de confianza respecto a la estimación puntual.

---oo0oo---

⌘ Ventajas

- Sabemos la probabilidad de que el valor verdadero se encuentre dentro del intervalo.
- Depende poco de la muestra escogida.
- Sabemos la estimación puntual sin más que tomar el centro del inter-valo y además será más fiable.

⌘ **Inconvenientes** :

- Hay que conocer (o estimar) la desviación típica de la población, lo que normalmente no sucede, pues presupondría saber la media a estimar
- Es más largo su cálculo y requiere una distribución de probabilidad conocida, o un tamaño muestral alto.

❁ ❁ ❁ ❁ ❁❁ ❁ ❁ ❁ ❁ ❁

❻ ¿Qué diferencia existe entre el nivel de confianza y el nivel de signi-ficación? ¿Es lógico hablar de un nivel de significación del 95%? ¿Y de un nivel de confianza del 95 %?

---oo0oo---

⌖ **Nivel de confianza** : Probabilidad (normalmente en porcentaje) de que un parámetro determinado pertenezca al inter-valo de confianza. En tanto por uno se le suele llamar grado de confianza ($1 - \alpha$).

⌖ **Nivel de significación** : lo contrario o complementario del nivel de confianza, la probabilidad de que un parámetro determi-nado no pertenezca al intervalo de confianza (en %) si se expresa en tanto por uno se le suele nominar grado de significación (α).

Un nivel de significación quiere decir que tenemos un 95 % de probabilidad de que el parámetro a estimar no va a pertenecer al intervalo calculado, ¿ qué sentido tiene ?.

El nivel de confianza si tiene sentido pues nos dice que, con una probabilidad del 95 % el parámetro estará en ese intervalo.

❁ ❁ ❁ ❁ ❁❁ ❁ ❁ ❁ ❁ ❁

❼ Si [25, 45] es el intervalo de confianza para cierto parámetro con un nivel de confianza dado, ¿cuál es el error máximo que se puede cometer, con ese mismo nivel de confianza?

---oo0oo---

Si tomamos como valor estimado del parámetro el valor medio del intervalo :

$$\hat{\theta} = \frac{25+45}{2} = 35$$

El error máximo que podemos cometer será la semiamplitud del intervalo :

$$E_{max} = \frac{54-25}{2} = 35 - 25 = 45 - 35 = 10.$$

❽ Al aumentar el nivel de confianza, ¿aumenta o disminuye la longitud del intervalo de confianza? ¿ Y al incrementarse el tamaño de la muestra? Razona tus respuestas.

---oo0oo---

Lógicamente, si deseamos aumentar la confianza de que un parámetro determinado este dentro de un intervalo, la amplitud de este intervalo ha de ser mayor para que comprenda más valores y tengamos más posibilidad de que pertenezca a él.

Como la desviación típica, en general es inversamente proporcional a la raíz cuadrada del tamaño de la muestra, al aumentar el tamaño de la muestra disminuye la desviación típica y al disminuir la desviación típica el tamaño del intervalo disminuye, está menos disperso entorno al valor central :

Si n⇧ como $\sigma \equiv 1/(n)^{0'5}$ ⇨ σ ⇩ y por tanto la amplitud el intervalo $(z_{\alpha/2} \cdot \sigma)$ ⇩

Ejercicios y problemas

❾ En una fábrica trabajan 387 mujeres y 478 hombres. Explica cómo seleccionar una muestra de 100 trabajadores usando muestreo aleatorio:

a) Simple.
b) Sistemático.
c) Estratificado proporcional al sexo.

---oo0oo---

a) Los 387 + 478 = 865 trabajadores se numeran (por orden de antigüedad, alfabético, o cualquier otro criterio) y se seleccionan 100 números por un procedimiento aleatorio.

b) Una vez numerados como en el caso anterior se selecciona al azar un número n_1 y el resto de los 99 se seleccionan a intervalo de :

$$k = \frac{865}{100} = 8'65$$

de 8 en 8, es decir :

n_1 , (n_1+8) , (n_1+16), (n_1+24), ... , $(n_1+8\cdot 98)$

c) Realizamos una afijación proporcional al tamaño de cada estrato para seleccionar los 100 trabajadores:

$$1^{er}\text{Estrato}: n_1 = \frac{387}{865}\cdot 100 = 44'75 \simeq 45$$

$$2^{o}\ \text{Estrato}: n_2 = \frac{478}{865}\cdot 100 = 55'26 \simeq 55$$

Luego seleccionamos **45 mujeres de las 397 y 55 hombres de los 478 mediante muestro aleatorio simple o sistemático.**

❿ Para conocer las preferencias musicales de los habitantes de una población, se decide entrevistar al 5 % de ellos usando muestreo aleatorio estratificado proporcional a la edad (niños-jóvenes-adultos). Si se ha calculado que esto supone entrevistar a 115 niños, 182 jóvenes y 398 adultos:

a) ¿Cuántos habitantes forman la población?
b) ¿Cuántos habitantes son niños? ¿Y jóvenes?

---oo0oo---

a) La muestra la componen :

n= $n_{niños} + n_{jóvenes} + n_{adultos}$ = 115 + 182 + 398 = = 695 individuos.

Como se han seleccionado un 5% de la población , el total de la población será :

$$N = \frac{695}{5} \cdot 100 = 13900 \text{ individuos.}$$

b) Como el muestreo es proporcional, la proporción en la muestra será la misma de la población :

$$N_{niños} = \frac{n_{niños}}{n} \cdot N = \frac{115}{695} \cdot 13900 = 2300$$

$$N_{jóvenes} = \frac{n_{jóvenes}}{n} \cdot N = \frac{182}{695} \cdot 13900 = 3640$$

❶❶ La estatura en centímetros de las mujeres mayores de edad de cierto país sigue una distribución N (158, 8). ¿Cuál es la probabilidad de que la estatura media de una muestra de tamaño 5 supere los 160 cm? ¿Y la de una muestra de tamaño 50? ¿Y la de una muestra de tamaño 500?
- Explica por qué la probabilidad disminuye al aumentar el tamaño de la muestra.

---oo0oo---

N(158, 8) ⇒ μ = 158 y σ = 8

a) n = 5
Distribución muestral :

$$\mu(\bar{X}) = \mu = 158, \sigma(\bar{x}) = \frac{\sigma}{\sqrt{n}} = \frac{8}{\sqrt{5}}$$

Aunque el tamaño es <30 la consideramos normal.

$$p(\bar{X} > 160) = p(z > \frac{160-158}{8/\sqrt{5}}) = p(z > 0'5590) =$$
= 1 - p(z ≤ 0’56) = 1 - 0’7123 = **0’2877**.

b) n = 50
Distribución muestral :

$$\mu(\bar{X}) = \mu = 158, \sigma(\bar{x}) = \frac{\sigma}{\sqrt{n}} = \frac{8}{\sqrt{50}}$$

Como el tamaño es > 30 la consideramos normal.

$$p(\bar{X} > 160) = p(z > \frac{160-158}{8/\sqrt{50}}) = p(z > 1'7678) =$$
= 1 - p(z ≤ 1’77) = 1 - 0’9616 = **0’0384**.

c) n = 500
Distribución muestral :

$$\mu(\bar{X}) = \mu = 158, \sigma(\bar{x}) = \frac{\sigma}{\sqrt{n}} = \frac{8}{\sqrt{500}}$$

Como el tamaño, n, es > 30 la consideramos normal.

$$p(\bar{X} > 160) = p(z > \frac{160-158}{8/\sqrt{500}}) = p(z > 5'590) =$$
= 1 - p(z ≤ 5’59) = 1 - 1 = **0**.

Observamos que al aumentar el tamaño de la muestra la media muestral se acerca cada vez más a la poblacional y por tanto la probabilidad de que la media de la muestra (160) sea distinta de la poblacional (158) va disminuyendo.

❶❷ La puntuación que obtienen los niños en cierto test psicológico sigue una distribución N (85, 15).

a) ¿Cuál es la probabilidad de que un niño elegido al azar obtenga más de 100 puntos?
b) ¿Cuál es la probabilidad de que la puntuación media en una muestra de 10 niños sea de más de 100 puntos?

---oo0oo---

Población : N(85, 15)

a) $p(X > 100) = p(z > \frac{100-\mu}{\sigma}) = p(z > \frac{100-85}{15}) =$

= p (z > 1) = 1 - p(z ≤ 1) = 1 - 0'8413 = **0'1517**

b) Ahora se trata de una muestra de tamaño n = 10, l distribución muestral se caracteriza por :

$$\mu(\overline{X}) = \mu = 85 \text{ y } \sigma(\overline{X}) = \frac{\sigma}{\sqrt{n}} = \frac{15}{\sqrt{10}}$$

que consideramos normal, aunque sería una t de Student.

$$p(\overline{X} > 100) = p\left(z > \frac{100-\mu(\overline{X})}{\sigma(\overline{X})}\right) = p\left(z > \frac{100-85}{15/\sqrt{10}}\right) =$$

= p (z > 3'16) = 1 - p (z ≤ 3'16) = 1 - 0'9992 = **0'0008**

❁ ❁ ❁ ❁ ❁❁ ❁ ❁ ❁ ❁ ❁

13 La nota de selectividad de los alumnos del centro A sigue una distribución N (6'5, 0'8), y la de los alumnos del centro B N (6'1, 1'4). Seleccionando al azar 45 alumnos de A y 52 alumnos de B, ¿cuál es la probabilidad de que la nota media de los alumnos de A supere en más de 0,25 puntos a la nota media de los alumnos de B?

---oo0oo---

- **Poblaciones** : A ≡ N (6'5, 0'8), B ≡ N(6'1,1'4).
- **Muestras** : n_A = 45 y n_B = 52 .
- **Distribución muestral** :
  - *Tipo* : de diferencia de medias.
  - *Media* : μ ($\bar{x}_A$ - $\bar{x}_B$) = μ_A - μ_B = 6'5 - 6'1 = 0'4.
  - *Desviación típica* :

$$\sigma(\bar{x}_A - \bar{x}_B) = \sqrt{\frac{\sigma_A^2}{n_A} + \frac{\sigma_B^2}{n_B}} = \sqrt{\frac{0'8^2}{45} + \frac{1'4^2}{52}} = 0'2278$$

Se nos pide hallar :

$$p(\bar{x}_A - \bar{x}_B > 0'25) = p\left(z > \frac{0'25-0'4}{0'2278}\right) = p(z > -0'66)$$

= p(z ≤ 0'66) = **0'7454** .

❁ ❁ ❁ ❁ ❁❁ ❁ ❁ ❁ ❁ ❁

14 El número de horas diarias que duermen los estudiantes de Bachillerato de una cierta comunidad autónoma se distribuye según una ley normal de media desconocida y desviación típica 3. A partir de una muestra de tamaño 30, se ha obtenido una media muestra¡ igual a 7 horas. Halla un intervalo de confianza al nivel del 96 % para la media de horas de sueño obtenida.

---oo0oo---

- **Población** : N(μ, 3).
- **Muestra** : n = 30, $\bar{x}$ = 7 horas.
- **Distribución muestral** :
  - *Tipo* : de medias.
  - *Desviación típica* :

$$\sigma(\bar{x}) = \frac{\sigma}{\sqrt{n}} = \frac{3}{\sqrt{30}} = 0'5477$$

- **Nivel de confianza** = 96 % ⇒ 1 - α = 0'96 ⇒ α = 0'04 ⇒ α/2 = 0'02 $z_{\alpha/2}$ = $z_{0'98}$ = 2'06 .

Luego el intervalo de confianza para la media poblacional es :

$$I.C. \equiv [\bar{x} \mp z_{\frac{\alpha}{2}} \cdot \sigma(\bar{x})] = [7 \mp 2'06 \cdot 0'5477] =$$

= [7 - 1'1283, 1 + 1'1283] = **[5'872, 8'128]**

❁ ❁ ❁ ❁ ❁❁ ❁ ❁ ❁ ❁ ❁

15 En las pruebas de selección de personal de una empresa, las puntuaciones obtenidas por los candidatos siguen una distribución N (l, 35). Sabiendo que en una muestra de 50 candidatos se observó una media de 75 puntos, halla el intervalo de confianza para la puntuación media correspondiente a los niveles de confianza del 99 %, el 95 % y el 90

%. Compara la amplitud de los intervalos obtenidos y extrae conclusiones.

---oo0oo---

Población : N(μ, 35).

Muestra : n = 50, $\bar{x}$ = 75 puntos.

Distribución muestral :

⌘ *Tipo* : de medias.

⌘ *Desviación típica* :

$$\sigma(\bar{x}) = \frac{\sigma}{\sqrt{n}} = \frac{35}{\sqrt{50}} = 4'9497$$

a) Nivel de confianza = 99 % ⇒ 1 - α = 0'99 ⇒ α = 0'01 ⇒ α/2 = 0'005 $z_{\alpha/2} = z_{0'995}$ = 2'58 .

Luego el intervalo de confianza para la media poblacional es :

$$[\bar{x} \mp z_{\frac{\alpha}{2}} \cdot \sigma(\bar{x})] = [75 \mp 2'58 \cdot 4'9497] =$$

= [75 - 12'77, 1 + 12'77] = **[62'23, 87'77]**

b) Nivel de confianza = 95 % ⇒ 1 - α = 0'95 ⇒ α = 0'05 ⇒ α/2 = 0'025 $z_{\alpha/2} = z_{0'975}$ = 1'96 .

Luego el intervalo de confianza para la media poblacional es :

$$[\bar{x} \mp z_{\frac{\alpha}{2}} \cdot \sigma(\bar{x})] = [75 \mp 1'96 \cdot 4'9497] =$$

= [75 - 9'7, 1 + 9'7] = **[65'3, 84'7]**

c) Nivel de confianza = 90 % ⇒ 1 - α = 0'9 ⇒ α = 0'1 ⇒ α/2 = 0'05 $z_{\alpha/2} = z_{0'95}$ = 1'65 .

Luego el intervalo de confianza para la media poblacional es :

$$[\bar{x} \mp z_{\frac{\alpha}{2}} \cdot \sigma(\bar{x})] = [75 \mp 1'65 \cdot 4'9497] =$$

= [75 - 8'167, 1 + 8'167] = **[66'833, 83'167]**

Se puede observar que a medida que disminuye el nivel de confianza el intervalo tiene menos amplitud.

❁ ❁ ❁ ❁ ❁❁ ❁ ❁ ❁ ❁ ❁

16 Para analizar el peso de unos botes de conserva, se toma una muestra de tamaño 32. Los pesos en kilogramos obtenidos son:

0'97	0'99	1'00	0'98	0'99	1'00	0'98	0'98
1'00	1'02	0'97	0'97	0'99	0'99	0'99	0'96
0'98	1'00	0'99	1'01	1'00	1'00	0'98	0'99
0'99	0'98	0'97	0'97	0'97	1'01	0'96	0'92

Halla el intervalo de confianza al nivel del 95% para el peso medio de los botes.

- Si el fabricante indica en los botes que el peso es de 1 kg, ¿es razonable pensar que está engañando a los clientes?

---oo0oo---

Población : Desconocida.

Muestra :

❖ Tamaño : n = 32 > 30

❖ Media y desviación típica, las hallamos mediante la tabla auxiliar:

x_i	f_i	$x_i \cdot f_i$	x_i^2	$x_i^2 \cdot f_i$
0'92	1	0'92	0'8464	0'8464
0'96	2	1'92	0'9216	1'8432
0'97	5	4'85	0'9409	4'7045
0'98	6	5'88	0'9604	5'7624
0'99	8	7'92	0'9801	7'8408
1'00	5	5	1	5
1'01	3	3'03	1'0201	3'0503
1'02	1	1'02	1'0404	1'0404
1'03	1	1'03	1'0609	1'0609
Suma	**32**	**31'57**		**31'1489**

$$\bar{x} = \frac{\sum_{i=1}^{32} x_i \cdot f_i}{n} = \frac{31'57}{32} = 0'98656$$

$$s = \sqrt{\frac{\sum_{i=1}^{32} x_i^2 \cdot f_i}{n} - \bar{x}^2} = \sqrt{\frac{31'1489}{32} - 0'9866^2} =$$

$$= 0'0202 \Rightarrow \sigma \simeq \hat{s} = \sqrt{\frac{n}{n-1}} \cdot S = 0'0206$$

❈ **Distribución muestral** :

⌘ *Tipo* : de medias.

⌘ *Desviación típica* :

$$\sigma(\bar{x}) = \frac{\sigma}{\sqrt{n}} = \frac{0'0206}{\sqrt{32}} = 0'0036$$

❈ **Nivel de confianza** = 95 % ⇒ 1 - α = 0'95 ⇒ α = 0'05 ⇒ α/2 = 0'025 $z_{\alpha/2} = z_{0'975}$ = 1'96 .

Luego el intervalo de confianza para la media poblacional es :

$$[\bar{x} \mp z_{\frac{\alpha}{2}} \cdot \sigma(\bar{x})] = [0'987 \mp 1'96 \cdot 0'0036] =$$

= [0'987 - 0'007, 0'987 + 0'007] = **[0'9799, 0'994]**

Como el peso indicado por el fabricante no pertenece al intervalo de confianza, con un riesgo del 5 %, podemos afirmar que el peso de los botes es menor de 1 kg.

❁ ❁ ❁ ❁ ❁❁ ❁ ❁ ❁ ❁ ❁

❶❼ Para estudiar la proporción de votos que obtendrá cierto partido político en unas elecciones, se hace un sondeo a 1 000 personas. De ellas, el 30% votará a ese partido. Suponiendo que se mantiene la intención de voto, halla el intervalo de confianza para la proporción de votos del partido, con un nivel de significación del 8%.
- Interpreta el resultado.

---oo0oo---

✪ **Muestra** :

❍ Tamaño : n = 1000 individuos.

❍ Proporciones : p = 0'3, q = 0'7.

✪ **Distribución muestral** :

❍ Tipo: de proporciones.

❍ Desviación típica :

$$\sigma(p) = \sqrt{\frac{p \cdot q}{n}} = \sqrt{\frac{0'3 \cdot 0'7}{1000}} = 0'014491$$

✪ **Nivel de significación** : α = 0'08 ⇒ α /2 = 0'04 ⇒ $z_{/2} = z_{0'96}$ = 1'76.

Luego el intervalo de confianza :

$$[p \mp z_{\frac{\alpha}{2}} \sigma(p)] = [0'3 \mp 1'76 \cdot 0'014491] =$$

= [0'3 - 0'0255, 0'3 + 0'0255] =

= [0'274, 0'326].

Significa que, con un 8 % de error, la proporción de votos del partido estará comprendida entre el 27'4 % y el 32'6 %.

❁ ❁ ❁ ❁ ❁❁ ❁ ❁ ❁ ❁ ❁

❶❽ Se desea comprobar la eficacia de dos tipos de somnífero en pacientes con insomnio. El somnífero A dio, en una muestra de 60 pacientes, una media de 7,15 h de sueño, con desviación típica de 0,65 h. El somnífero B dio, en una muestra de 80 pacientes, una media de 6,85 h de sueño, con desviación típica de 1,15 h. Halla el intervalo de confianza para la diferencia de medias, con un nivel de significación del 5 %.

- ¿Es razonable aceptar que ambos somníferos son igual de eficaces? Razona tu respuesta.

---oo0oo---

❅ **Muestra A** : $n_A = 60$, $\bar{x}_A = 7'15$, $s_A = 0'65$.

❅ **Muestra B** : $n_B = 80$, $\bar{x}_B = 6'85$, $s_B = 1'15$

❅ **Nivel de significación** : α = 0'05 ⇒ α /2 = 0'025 ⇒ $z_{/2} = z_{0'975}$ = 1'96

❅ **Distribución Muestral** :

❒ Tipo : de diferencia de medias.

❒ Desviación típica :

$$\sigma(\bar{x}_A - \bar{x}_B) = \sqrt{\frac{\frac{n_A}{n_A-1}s_A^2}{n_A} + \frac{\frac{n_B}{n_B-1}s_B^2}{n_B}} =$$

$$= \sqrt{\frac{s_A^2}{n_A-1} + \frac{s_B^2}{n_B-1}} = \sqrt{\frac{0'65^2}{60-1} + \frac{1'15^2}{80-1}} = 0'1546$$

ya que, al no conocer la desviación típica de las poblaciones y ser los tamaños mayores que 30 usamos las desviaciones muestrales.

El intervalo de confianza para la diferencia de medias poblacionales es :

$$\left[(\bar{x}_A - \bar{x}_B) \mp z_{\frac{\alpha}{2}} \cdot \sigma(\bar{x}_A - \bar{x}_B)\right] =$$

= [(7'15 - 6'85) ∓ 1'96 · 0'1546] =

= [0'3 ∓ 0'3030] = **[- 0'003, 0'603]**

Si son igual de eficaces las medias han de ser iguales y , por tanto, su diferencia nula. Como esta diferencia de medias está incluida en el intervalo de confianza, podemos concluir que sí pueden ser igual de eficaces.

❋ ❋ ❋ ❋ ❋❋ ❋ ❋ ❋ ❋ ❋

❶❾ Comprueba que los siguientes datos, obtenidos de una ficha técnica, son correctos:

Muestra: 2500 entrevistas, con un error posible de -2 %, para un nivel de confianza del 95,5 % (dos sigma) p/q = 50/50.

- ¿Cuál debería ser el tamaño de la muestra para tener, con una confianza del 90%, un error posible E = ±3 %?

---oo0oo---

❍ Tamaño : n = 2 500 individuos.

❍ Error máximo E = ± 2 %.

❍ **Nivel de confianza** = 95'5 % ⇒ 1 - α = 0'955 ⇒ α = 0'045 ⇒ α/2 = 0'0225 $z_{\alpha/2} = z_{0'9775}$ = 2'01 .

❍ p/q= 50/50 ⇒ p = q = 0'5

Comprobemos el error máximo:

$$E = z_{\frac{\alpha}{2}} \cdot \sqrt{\frac{p \cdot q}{n}} = 2'01\sqrt{\frac{0'5 \cdot 0'5}{2500}} \simeq 0'02 \ (2\ \%)$$

Luego los datos son correctos.

---oo0oo---

❍ **Nivel de confianza** = 90 % ⇒ 1 - α = 0'9 ⇒ α = 0'1 ⇒ α/2 = 0'05 $z_{\alpha/2} = z_{0'95}$ = 1'65 .

Aplicando la fórmula del tamaño :

$$n = \frac{z_{\frac{\alpha}{2}}^2 \cdot p \cdot q}{E^2} = \frac{1'65^2 \cdot 0'5^2}{0'03^2} = 756'25$$

Hemos de tomar una muestra de 757 individuos para asegurarnos un error máximo del 3 %.

❋ ❋ ❋ ❋ ❋❋ ❋ ❋ ❋ ❋ ❋

❷⓿ El peso en gramos de ciertas magdalenas tiene distribución N (30, 5). Si esas magdalenas se envasan en bolsas de 16 unidades:

a) Halla la probabilidad de que, eligiendo una bolsa al azar, su contenido supere los 468 g.

b) El contenido del 67% de las bolsas no supera cierto peso x. ¿De qué peso x se trata?

---oo0oo---

➠ **Población** : N (30, 5).

➠ **Muestra** : n = 16.

➠ **Distribución muestral** :

⦿ *Tipo* : de medias.

⦿ *Media* : $\mu(\bar{x}) = \mu = 30$.

⦿ *Desviación típica* :

$$\sigma(\bar{x}) = \frac{\sigma}{\sqrt{n}} = \frac{5}{\sqrt{16}} = \frac{5}{4} = 1'25$$

Luego la distribución muestral es una :

N(30, 1’25)

a) Se nos pide hallar :

$$p(\bar{x} > \frac{468}{16}) = p(\bar{x} > 29'25) = p\left(z > \frac{29'25-30}{1'25}\right)$$

= p (z > -0’6) = p(z ≤ 0’6) = Φ(0’6) = **0’7257**.

b) Primero hallamos el valor de z que deja bojo de si un 67 % :

p(Z ≤ z) = 0’67 , buscando en la tabla de la N(0, 1) vemos que z = 0’44.

Ahora en la distribución muestral de medias despejamos la media que corresponde a ese z :

$$z = \frac{\bar{x}-\mu(\bar{x})}{\sigma(\bar{x})}; \frac{\bar{x}-30}{1'25} = 0'44; \bar{x} = 30 + 0'44 \cdot 1'25$$

$$\bar{x} = 30'55 \Rightarrow \frac{x}{16} = 30'55 \Leftrightarrow x = 30'55 \cdot 16$$

x = 488’8 g.

❁ ❁ ❁ ❁ ❁❁ ❁ ❁ ❁ ❁ ❁

❷❶ Se desea estudiar la edad media en años de los espectadores que asisten a cierta proyección cinematográfica. En una muestra de 50 espectadores se observa:

38	41	25	62	14	5	33	49	7	13
29	52	61	73	32	37	34	45	43	57
65	18	23	80	68	60	54	38	46	55
12	24	14	39	31	44	58	47	36	37
26	25	19	34	33	42	53	41	56	59

a) Halla el intervalo de confianza para la edad media de los espectadores, con nivel de significación del 1 %, y obtén el error máximo.

b) ¿A cuántos espectadores más se tendría que encuestar para que el error máximo cometido fuera de 2 años?

---ooOoo---

❈ **Población** : Desconocida.

❈ **Muestra** :

❖ Tamaño : n = 50 > 30

❖ Media y desviación típica, las hallamos la mediante la calculadora :

$\bar{x}$ = 39’74 ; s_{n-1} = 17’71637

❈ **Distribución muestral** :

⌘ *Tipo* : de medias.

⌘ *Desviación típica* :

$$\sigma(\bar{x}) = \frac{s_{n-1}}{\sqrt{n}} = \frac{17'71637}{\sqrt{50}} = 2'5055$$

❈ **Nivel de** significación = 1 % ⇒ α = 0’01 ⇒ α/2 = 0’005 $z_{\alpha/2} = z_{0'995}$ = 2’58 .

Luego el intervalo de confianza para la media poblacional es :

$$[\bar{x} \mp z_{\frac{\alpha}{2}} \cdot \sigma(\bar{x})] = [39'74 \mp 2'58 \cdot 2'5055] =$$

= [39’74 - 6’464, 39’74 + 6’464] =

= [33’2758, 46’204]

El error máximo :

$$E = z_{\frac{\alpha}{2}} \cdot \sigma(\bar{x}) = 6'464 \text{ años}$$

b)

$$n = \left(\frac{z \cdot \sigma}{E}\right)^2 = \left(\frac{2'58 \cdot 17'71637}{2}\right)^2 = 522'31$$

Para que el error no supere el 2% hemos de entrevistar a 523 personas, como ya hemos entrevistado a 50 nos quedan por encuestar : 523 - 50 = 473 individuos.

❷❷ Un novelista desea conocer el porcentaje de habitantes de su ciudad que comprarán su próxima obra. Para ello, encarga a una empresa que entreviste a 100 habitantes. Sabiendo que 32 de ellos manifiestan querer adquirir la obra, averigua cuál es el error máximo cometido en la predicción y a cuántos habitantes más se ha de entrevistar para que, con nivel de confianza del 90%, el error máximo sea del 5%.

---oo0oo---

➢ **Muestra** :

- *Tamaño* : n = 100.
- *Proporción* = p = 32/100 = 0'32 , q = 1-p = 1 - 0'32 = 0' 68.

➢ **Nivel de confianza** = 90 % ⇒ α = 0'1 ⇒ α/2 = 0'05 $z_{\alpha/2} = z_{0'95}$ = 1'65.

➢ **Error muestral** :

$$E = z_{\frac{\alpha}{2}} \sqrt{\frac{p \cdot q}{n}} = 1'65 \cdot \sqrt{\frac{0'32 \cdot 0'68}{100}} = 0'769$$

Es decir un error del 7'7 % .

Si el error no debe superar el 5 %, se ha de entrevistar a :

$$n = \frac{z^2 \cdot p \cdot q}{E^2} = \frac{1'65^2 \cdot 0'32 \cdot 0'68}{0'05^2} = 236'96$$

es decir 237 individuos, **como se han encuestado 100, se habrá de encuestar a 237 - 100 = 137 individuos más.**

❷❸ El tiempo en minutos que tardan en reparar cierta avería en un taller A sigue una distribución N (μ_1, 25). En otro taller B, dicho tiempo de reparación sigue una distribución N (μ_2, 30). En una muestra de 10 reparaciones de ese tipo de avería en el taller A, se observó un tiempo medio de 80 min. mientras que en una muestra de 15 reparaciones en B la media fue de 75 min.

a) Halla el intervalo de confianza para la diferencia de tiempos medios, con un nivel de significación del 1 %, y di cuál es el error máximo cometido al hacer dicha estimación.

b) Si tomamos muestras de igual tamaño en ambos talleres, ¿cuál ha de ser este tamaño para conseguir un error máximo de 10 min?

---oo0oo---

a)

▣ **Poblaciones** : N (μ_1, 25) y N (μ_2, 30).

▣ **Muestras** : $n_1 = 10$ $\bar{x}_1 = 80$, $n_2 = 15$, $\bar{x}_2 = 75$

▣ **Distribución muestral** :

- *Tipo* : de diferencia de medias
- *Desviación típica* :

$$\sigma_{\bar{x}_1 - \bar{x}_2} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} = \sqrt{\frac{25^2}{10} + \frac{30^2}{15}} = 11'07$$

▣ **Nivel de significación** = 1 % ⇒ α = 0'01 ⇒ α/2 = 0'005 $z_{\alpha/2} = z_{0'995}$ = 2'58.

▣ **Intervalo de confianza** para la diferencia de medias poblacionales :

$$\left[(\bar{x}_1 - \bar{x}_2) \mp z_{\frac{\alpha}{2}} \cdot \sigma_{\bar{x}_1 - \bar{x}_2}\right] =$$

$$[(80 - 75) \mp 2'58 \cdot 11'07] = [5 \mp 28'561] =$$

= [-23'56, 33'56]

▣ **Error máximo**:

$$E = z_{\frac{\alpha}{2}} \cdot \sigma_{\bar{x}_1 - \bar{x}_2} = 28'56 \text{ minutos}$$

b) $n_1 = n_2 = n$

$$E = 10 = z_{\frac{\alpha}{2}} \cdot \sigma_{\bar{x}_1 - \bar{x}_2} = 2'58 \cdot \sqrt{\frac{25^2}{n} + \frac{30^2}{n}} \Rightarrow$$

$$2'58 \cdot \sqrt{\frac{25^2}{n}+\frac{30^2}{n}} = 10; \sqrt{\frac{625+900}{n}} = 3'876$$

$$n = \frac{1525}{3'876^2} = 101'51$$

Es decir 102 individuos.

❷❹ Lanzamos una moneda perfecta 100 veces.

a) Halla la probabilidad de que el número de caras no esté comprendido entre 46 y 54.
b) Halla L tal que [50 - L, 50 + L] sea el intervalo de confianza para el número de caras, al nivel de significación del 5 %.

---oo0oo---

a)

◈ **Población** : P = Q = 0'5 (en una moneda perfecta las dos posibilidades son iguales).

◈ **Muestra** = n = 100 p_1 = 0'46 p_2 = 0'54.

◈ **Distribución muestral**:

▣ *Tipo* : de proporciones.
▣ *Media* $\mu(p)$ = P = 0'5
▣ *Desviación típica* :

$$\sigma(p) = \sqrt{\frac{P \cdot Q}{n}} = \sqrt{\frac{0'5 \cdot 0'5}{100}} = 0'05$$

▣ Como n >30 es una N(0'5 , 0'05)

◈ **Probabilidad** de que la muestra pertenezca al intervalo [46/100, 54/100] :

$$p(0'46 \le p \le 0'54) =$$

$$= p\left(\frac{0'46-\mu(p)}{\sigma(p)} \le p \le \frac{0'54-\mu(p)}{\sigma(p)}\right) =$$

$$= p\left(\frac{0'46-0'5}{0'05} \le p \le \frac{0'54-0'5}{0'05}\right) =$$

= p (- 0'8 ≤ z ≤ 0'8) = p(z ≤ 0'8) - p(z ≤ - 0'8) = p(z ≤ 0'8) - [1- p(z ≤ 0'8)] = 2·p(z ≤ 0'8) -1 = 2 ·Φ(0'8) - 1 = 2 · 0'7871 - 1 = 0'5742.

Como la probabilidad pedida es que no pertenezca a dicho intervalo :

1 - 0'5742 = 0'4258

b)

El intervalo lo dividimos por 100 para pasar a proporciones y queda :

$$\left[\frac{50-L}{100}, \frac{50+L}{100}\right] = [0'5 - 0'01L,\ 0'5 + 0'01\ L\ ]$$

cuyo error máximo es 0'01 L, e igualando a la fórmula del error máximo :

$$E = 0'01L = z_{\frac{\alpha}{2}} \cdot \sigma(p) = 1'96 \cdot 0'05$$

ya que con α = 0'05 z = 1'96. Luego :

0'01L = 0'098 ; **L= 0'098/0'01 = 9'8**.

❷❺ Sabiendo que, si $n_1 \ge 30$ y $n_2 \ge 30$, la distribución de muestreo de la diferencia de proporciones es:

$$N\left(p_1 - p_2, \sqrt{\frac{p_1 \cdot (1-p_1)}{n_1} + \frac{p_2 \cdot (1-p_2)}{n_2}}\right)$$

indica la forma general del intervalo de confianza para la diferencia de proporciones.

- En una muestra de 150 jóvenes y 200 adultos, a 78 jóvenes y a 84 adultos les gusta cierto grupo musical. Halla el intervalo de confianza para la diferencia de proporciones con nivel de significación del 8 %.

---oo0oo---

⦿ Análogamente al intervalo de confianza de la diferencia de medias , el intervalo de diferencia de proporciones es :

$$\left[(p_1 - p_2) \mp z_{\frac{\alpha}{2}} \sqrt{\frac{P_1 \cdot Q_1}{n_1} + \frac{P_2 \cdot Q_2}{n_2}}\ \right]$$

en donde p_i = proporciones de las muestras, y P_i = proporciones de las poblaciones.

⦿ ▣ **Poblaciones** : P_i = desconocidas

▣ **Muestras** : $n_1 = 150$, $n_2 = 200$, $p_1 = 78/150 = 0'52$, $p_2 = 84/200 = 0'42$.

▣ **Distribución muestral** :

✦ *Tipo* : de diferencia de proporciones.

✦ *Media* : $\mu_{p1-p2} = P_1 - P_2 \approx p_1 - p_2 = 0'52 - 0'42 = 0'1$

✦ *Desviación típica* :

$$\sigma_{p_1-p_2} = \sqrt{\frac{P_1 \cdot Q_1}{n_1} + \frac{P_2 \cdot Q_2}{n_2}} \simeq \sqrt{\frac{p_1 \cdot q_1}{n_1} + \frac{p_2 \cdot q_2}{n_2}} =$$

$$= \sqrt{\frac{0'52 \cdot 0'48}{150} + \frac{0'42 \cdot 0'58}{200}} = 0'0537$$

Luego es una N(0'1, 0'0537)

▣ **Nivel de significación** = 8 % $\Rightarrow \alpha = 0'08 \Rightarrow \alpha/2 = 0'04$ $z_{\alpha/2} = z_{0'96} = 1'76$

▣ **El intervalo de confianza** será pues :

[0'1 1'76 · 0'0537] = **[0'0055, 0'1944]**