AUTOEVALUACIÓN DE MUESTREO E INFERENCIA ESTADÍSTICA

Sin que aparezcan en este orden y en algún caso combinados, estos son los tipos básicos de los problemas que se plantean: ***1)*** *Tipos de muestreos.* ***2)*** *Cálculo de probabilidades de una variable aleatoria con distribución normal.* ***3)*** *Cálculo de probabilidades de la media muestral.* ***4)*** *Cálculo de probabilidades de la proporción muestral.* ***5)*** *Cálculo de probabilidades de la diferencia de medias muestrales.* ***6)*** *Estimación puntual y por intervalos de confianza para la media muestral.* ***7)*** *Estimación puntual y por intervalos de confianza para la proporción muestral.* ***8)*** *Estimación puntual y por intervalos de confianza para la diferencia de medias muestrales.* ***9)*** *Cálculo del error máximo admisible de una estimación.* ***10)*** *Cálculo del mínimo tamaño muestral que acote el error máximo admisible de una estimación.* ***11)*** *Contraste de hipótesis unilateral o bilateral para la media.* ***12)*** *Contraste de hipótesis unilateral o bilateral para la proporción.* ***13)*** *Contraste de hipótesis unilateral o bilateral para la diferencia de medias.*

1. En cierto país, la talla de los recién nacidos se distribuye normalmente con una media de 66 cm. y una desviación típica de 5 cm.
a) ¿Qué porcentaje de recién nacidos cabe esperar que midan entre 65 y 70 cm?
b) De los 130 que nacieron este mes, ¿cuántos de ellos medirán entre 65 y 70 cm?

2. Un estudio realizado sobre 100 usuarios revela que sus automóviles recorren anualmente un promedio de 15 200 km. Suponiendo una distribución normal con desviación típica de 2 250 km:
a) Determine un intervalo de confianza, al 97%, para la cantidad promedio de kilómetros recorridos.
b) ¿Cuál sería el máximo error cometido en esta estimación?
c) ¿Cuál debe ser el tamaño mínimo de la muestra para que el error cometido no sea superior a 250 km, con igual confianza?

3. Cierto día, un supermercado toma nota de la cantidad de dinero gastado por algunos de sus clientes. Los resultados son los siguientes:

Dinero gastado (en €)	0 - 20	20 - 40	40 - 60	60 - 80	80 - 100	100 - 120	120 – 140
Número de clientes	24	16	22	40	18	10	4

a) Haz una estimación de la cantidad media que gastan habitualmente los clientes de dicho supermercado.
b) Comenta cómo modificarías el tipo de muestreo realizado con el fin de mejorar su representatividad.

4. El peso de los adultos de una población numerosa se distribuye normalmente con una media de 65 Kg y 12 Kg de desviación típica: Se elige una muestra de 36 individuos al azar. Calcula la probabilidad de que el peso medio de dicha muestra:
a) sea mayor de 60 Kg;
b) esté en el intervalo (60, 64).

5. Los datos recopilados durante varias promociones aseguran que aproximadamente el 90% de los alumnos de 2º de Bachillerato del IES Jovellanos de Gijón aprueban el examen de Matemáticas de selectividad. En un estudio hecho por el profesor de matemáticas, en la selectividad del 2011, aprobaron el examen de Matemáticas 76 alumnos de los 78 presentados. El profesor defiende que el porcentaje de aprobados ha aumentado.
a) Con un nivel de significación del 5%. ¿Puede defenderse que el porcentaje de aprobados no ha aumentado? Plantear el contraste o test de hipótesis y resolverlo.
b) A la luz de los nuevos datos obtenidos, determinar un intervalo de confianza al 95% del porcentaje de aprobados de los alumnos de 2º de Bachillerato del IES Jovellanos de Gijón.

6. Un taller está especializado en la fabricación de cojinetes de los que se usan en pequeñas máquinas. Las empresas que los incorporan en sus productos exigen que estos cojinetes tengan un diámetro exacto, ya que si fuese mayor causaría vibraciones en el eje que se apoya en ellos y si fuese menor produciría rozamientos, alterando en ambos casos el buen funcionamiento de la maquinaria.
El taller ha conseguido que sus cojinetes dispongan de un diámetro medio de 0,4 mm con una pequeña desviación típica de 0,01 mm.
En el último estudio de calidad realizado en el taller, encontraron que, en un muestreo de 200 cojinetes, el diámetro medio fue de 0,402 mm, resultado que les preocupó.
Plantea un test para contrastar la hipótesis de que se mantiene la misma precisión en la fabricación de estos cojinetes frente a que, por alguna razón que se tendría que estudiar, resulten menos precisos.

¿A qué conclusión se llega con un nivel de significación del 1%?

7. Al acto de presentación de unas oposiciones asistió el 65% de los candidatos. Si se hubiesen tomado, elegidos al azar, 81 opositores, ¿cuál es la probabilidad de que, de ellos, se presenten menos de 55?

8. Se sabe que tanto para niños como para niñas de sexto de primaria los pesos siguen una distribución normal. El peso medio de los niños de sexto de primaria es de 45 Kg y su desviación típica es de 6,5 Kg, mientras que el promedio de los pesos de las niñas de ese mismo nivel educativo es de 40 Kg y 5,5 Kg de desviación típica. Realizado un muestreo aleatorio de 20 niños y 25 niñas de sexto de primaria, ¿cuál es la probabilidad de que el promedio de los pesos de esos 20 niños supere en al menos 8 Kg al de las 25 niñas?

9. En una universidad se toma al azar una muestra de 80 alumnos y se encuentra que 60 han aprobado todas las asignaturas.
a) Estima el porcentaje de aprobados en dicha universidad.
b) Con un nivel de confianza del 96%, halla un intervalo para estimar el porcentaje de alumnos que aprueban todas las asignaturas.
c) A la vista del resultado anterior se pretende repetir la experiencia para conseguir una cota de error del 5%, con el mismo nivel de confianza. ¿Cuántos individuos ha de tener la muestra?

10. Se comparan dos tipos de rosca de tornillo para ver su resistencia a la tensión. Se prueban 50 piezas de cada tipo bajo condiciones similares, la marca A tuvo una resistencia promedio a la tensión de 78.3 Kg, mientras que la marca B tuvo una resistencia promedio de 87.2 Kg. Se sabe de antemano que las desviaciones típicas son de 6.5 Kg para la marca A y 6.3 Kg para la B. Determine un intervalo de confianza al 93% para la diferencia de resistencia media a la tensión de las dos marcas de roscas.

11. El contenido de leche en las botellas llenadas por cierta máquina envasadora, antes de averiarse, se distribuía según una variable aleatoria normal de media 1000 cm^3 y desviación típica 20 cm^3.
Tras la reparación de la avería, la distribución de los contenidos de las botellas envasadas por la máquina sigue siendo normal con desviación típica de 20 cm^3, pero al tomar una muestra de 25 botellas llenadas por la máquina reparada se obtiene una media muestral de sus contenidos de 1005 cm^3.
a) Determine si se debe aceptar la hipótesis de que la media de los volúmenes envasados por la máquina tras la reparación sigue siendo como antes, o rechazarla a favor de que la media ha aumentado, con un nivel de significación del 5%.
b) Explicar, en el contexto del problema, en que consisten cada uno de los errores tipo I y II.

12. Supongamos que en un centro escolar los alumnos se distribuyen de acuerdo con la tabla siguiente:

3º ESO	4º ESO	1º Bach	2º Bach
95	106	108	91

a) Si la variable estadística a estudiar fuese el número de personas que viven en la misma casa que estos estudiantes, ¿sería preferible un muestreo estratificado por nivel de estudios a uno aleatorio simple sin estratificar?
b) ¿Y qué muestreo preferiríamos si la variable estadística a estudiar fuese la hora en que llegan a casa los sábados?
c) Si por fin se decidiera elegir a 75 alumnos utilizando el método de muestreo estratificado por nivel de estudios, ¿a cuantos estudiantes de cada tipo deberás encuestar?

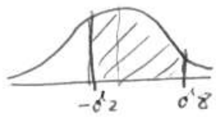
13. Deseamos conocer si los vinos con la denominación de origen de Ribera de Duero tienen el mismo contenido alcohólico que los de Toro ya que, debido a la proximidad geográfica de ambas regiones, es posible que haya fraudes y se intercambien vinos de ambas denominaciones dependiendo del mercado de los mismos. Supondremos que en ambos casos se sigue una distribución normal con desviaciones típicas conocidas: 0,5 y 0,6 para las denominaciones de Ribera de Duero y Toro respectivamente. La graduación media de alcohol obtenida en una muestra aleatoria de 14 vinos de Ribera de Duero es de 12,529º mientras que la obtenida en una muestra aleatoria de 6 vinos de Toro es de 13,450. Diseña un test para contrastar la hipótesis de que los dos tipos de vino tienen la misma graduación. ¿A qué conclusión se llega con un nivel de significación del 2%?

① $N(66,5)$

X: Talla de los recién nacidos.

a) $P[65<X<70] = P\left[\frac{65-66}{5} < Z < \frac{70-66}{5}\right] = P[-0'2<Z<0'8] =$

$= 0'7881-(1-0'5793) = 0'3674 = \boxed{36,74\%}$



b) $130 \cdot 36,74\% = 47,762 \rightarrow$ $\boxed{\text{48 de los 130 recién nacidos medirán entre 65 y 70}}$

② $N(\mu; 2250)$

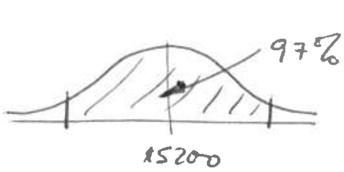
X: Kilómetros recorridos por los automóviles.

$\bar{x} = 15200$

$\sigma = 2250$

$n = 100$

$1-\alpha = 97\% \Rightarrow Z_{\alpha/2} = 2'17$



$\left(15200 - 2'17 \cdot \frac{2250}{\sqrt{100}} ; 15200 + 2'17 \cdot \frac{2250}{\sqrt{100}}\right)$

$\boxed{\mu \in (14711'75; 15688'25) \text{ con una confianza del } 97\%}$

b) Error Máximo $= 2'17 \cdot \frac{2250}{\sqrt{100}} = \boxed{488,25 \text{ Km}}$

c) $250 = 2'17 \cdot \frac{2250}{\sqrt{n}} \Rightarrow n = \left(\frac{2'17 \cdot 2250}{250}\right)^2 = 381'42 \rightarrow \boxed{n = 382}$

③ $\bar{X} = \frac{\sum x_i f_i}{\sum f_i} = \frac{10\cdot 24+30\cdot 16+50\cdot 22+70\cdot 40+90\cdot 18+110\cdot 10+130\cdot 4}{24+16+22+40+18+10+4} = \frac{7860}{134} = 58,66$ €

a) El gasto medio es $\boxed{58,66 \text{ €}}$

b) Podría ser más representativo si lo hiciésemos en días distintos durante la semana y en fines de semana y a distintas horas.

④ $N(65; 12)$

X: peso en Kg de los adultos en dicha población

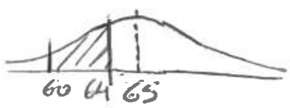
$n = 36$

$\bar{X}$ será $N\left(65; \frac{12}{\sqrt{36}}\right) = N(65; 2)$

a) $P[\bar{X} > 60] = P\left[Z > \frac{60-65}{2}\right] = P[Z > -2'5] = \boxed{0'9938}$

b) $P[60 < \bar{X} < 64] = P\left[\frac{60-65}{2} < Z < \frac{64-65}{2}\right] = P[-2'5 < Z < -0'5] =$

$= P[0'5 < Z < 2'5] = 0'9938 - 0'6915 = \boxed{0'3023}$



⑤ a)

$H_0: p \le 0'90$

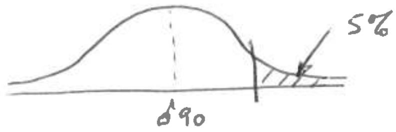
$H_1: p > 0'90$

Siendo p la proporción de alumnos que aprueban matemáticas

$\alpha = 5\% \Rightarrow z_\alpha = 1'645$

$\hat{p} = \frac{76}{78} = 0'974$

$n = 78$



$$\left(-\infty,\ 0'90 + 1'645 \cdot \sqrt{\frac{0'90 \cdot 0'10}{78}}\right) = (-\infty,\ 0'9559)$$

$0'974 \notin (-\infty, 0'9559)$ Valor significativo

Por lo tanto, rechazamos la hipótesis nula, es decir: con una significación del 5%, pensamos que la proporción de aprobados sí ha aumentado.

b)

$\hat{p} = 0'974$

$n = 78$

$1-\alpha = 95\% \Rightarrow z_{\alpha/2} = 1'96$

$$\left(0'974 - 1'96\sqrt{\frac{0'974 \cdot 0'026}{78}},\ 0'974 + 1'96\sqrt{\frac{0'974 \cdot 0'026}{78}}\right)$$

$\boxed{P_0 \in (0'9393;\ 1)}$ intervalo de confianza al 95% de la proporción de aprobados

⑥ a)

$H_0: \mu = 0'4$

$H_1: \mu \ne 0'4$

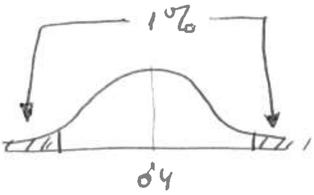
Siendo μ el diámetro de los cojinetes

$\alpha = 1\% \Rightarrow z_{\alpha/2} = 2'58$

$\bar{x} = 0'402$

$n = 200$

$\sigma = 0'01$



$$\left(0'4 - 2'58\frac{0'01}{\sqrt{200}},\ 0'4 + 2'58 \cdot \frac{0'01}{\sqrt{200}}\right) = (0'3982;\ 0'4018)$$

$0'402 \notin (0'3982;\ 0'4018)$ Valor significativo

Por lo tanto, con una significación del 1%, rechazamos la hipótesis nula, pensamos que el diámetro medio ya no es de 0'4 mm.

(7) $p_0 = 0'65$

$n = 81$

$\hat{p}$ es una $N\left(0'65 ; \sqrt{\frac{0'65 \cdot 0'35}{81}}\right) = N(0'65 ; 0'053)$

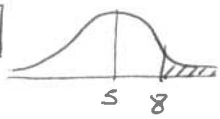
$P\left[\hat{p} < \frac{55}{81}\right] = P\left[Z < \frac{\frac{55}{81} - 0'65}{0'053}\right] = P[Z < 0'55] = \boxed{0'7088}$

(8) Niños : $N(45 ; 6'5)$ $\quad X_1$: peso de los niños ; $n_1 = 20$

Niñas : $N(40 ; 5'5)$ $\quad X_2$: " " las niñas ; $n_2 = 25$

$\bar{X}_1 - \bar{X}_2$ es una $N\left(45-40 ; \sqrt{\frac{6'5^2}{20} + \frac{5'5^2}{25}}\right) = N(5 ; 1'8228)$

$P[\bar{X}_1 - \bar{X}_2 > 8] = P\left[Z > \frac{8-5}{1'8228}\right] = P[Z > 1'65] = 1 - 0'9505 = \boxed{0'0495}$



(9) P : porcentaje de alumnos de la universidad que aprueban todas.

$\hat{P} = \frac{60}{80} = 0'75$

a) Estimamos que el porcentaje de aprobados es del $\boxed{75\%}$

b) $1-\alpha = 96\% \rightarrow Z_{\alpha/2} = 2'055$; $n = 80$

$\left(0'75 - 2'055\sqrt{\frac{0'75 \cdot 0'25}{80}} ; 0'75 + 2'055\sqrt{\frac{0'75 \cdot 0'25}{80}}\right) = (0'6505 ; 0'8495)$

Al 96%, el porcentaje de alumnos que aprueban todas las asignaturas estará en el intervalo $\boxed{(65'05\% ; 84'95\%)}$

c) $0'03 = 2'055 \cdot \sqrt{\frac{0'75 \cdot 0'25}{n}} \Rightarrow n = \frac{0'75 \cdot 0'25}{\left(\frac{0'03}{2'055}\right)^2} = 316'73 \Rightarrow \boxed{n = 317}$

(10) X_A : Resistencia de la marca A

X_B : " " " " B

$\bar{X}_A = 78'3$

$n_A = 50$

$\sigma_A = 6'5$

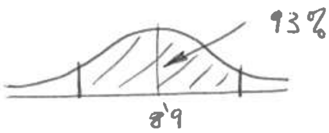
$\bar{X}_B = 87'2$

$n_B = 50$

$\sigma_B = 6'3$

$1-\alpha = 93\% \rightarrow Z_{\alpha/2} = 1'81$

$\mu_B - \mu_A$ es una $N\left(87'2 - 78'3 ; \sqrt{\frac{6'5^2}{50} + \frac{6'3^2}{50}}\right) = N(8'9 ; 1'28)$



$(8'9 - 1'81 \cdot 1'28 ; 8'9 + 1'81 \cdot 1'28)$

$\boxed{\mu_B - \mu_A \in (6'58 ; 11'22)}$ al 93% de confianza

(11) a) $H_0: \mu \leq 1000$ Siendo μ el contenido medio de leche en las botellas.

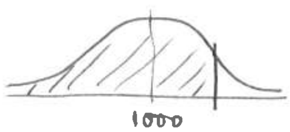
$H_1: \mu > 1000$

$n = 25$

$\sigma = 20$

$\bar{x} = 1005$

$\alpha = 5\% \rightarrow z_\alpha = 1'645$



$$\left(-\infty, 1000 + 1'645 \cdot \frac{20}{\sqrt{25}}\right) = (-\infty, 1006'58)$$

$1005 \in (-\infty, 1006'58)$, valor no significativo.

Por lo tanto, con una significación del 5%, pensamos que la máquina sigue como antes de la avería y el contenido medio de leche no ha aumentado.

b)

Ho cierta — Aceptamos Ho

Ho cierta — Error I — Rechazamos Ho

Ho falsa — Error II — Aceptamos Ho

Ho falsa — Rechazamos Ho

Error Tipo I: pensar que el contenido medio sí ha aumentado cuando en realidad sigue igual.

Error Tipo II: Pensar que el contenido medio sigue igual que antes cuando en realidad ha aumentado.

(12) a) Como dicha variable no parece ser relevante en el nivel de estudios, se podría hacer un muestreo aleatorio juntando a todos los alumnos.

b) Si la variable a estudiar se prevé que pudiera ser sustancialmente distinta para diferentes edades, como es el caso, sería preferible hacer estratos por niveles educativos.

c) En total son: $95+106+108+91 = 400$ alumnos.

$95 \cdot \frac{75}{400} = 17'81$; $106 \frac{75}{400} = 19'88$; $108 \cdot \frac{75}{400} = 20'25$; $91 \cdot \frac{75}{400} = 17'06$

Redondeando: $\boxed{\text{18 de 3º ESO, 20 de 4º ESO, 20 de 1º Bach., 17 de 2º Bach}}$

(13) $H_0: \mu_1 - \mu_2 = 0$
$H_1: \mu_1 - \mu_2 \neq 0$

Siendo μ_1 la graduación media de alcohol de los vinos Ribera y μ_2 " " " " " " " de Toro.

$n_1 = 14$

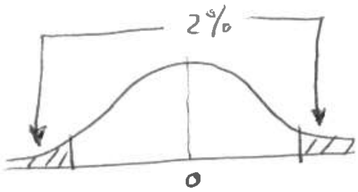
$n_2 = 6$

$\sigma_1 = 0'5$

$\sigma_2 = 0'6$

$\bar{x}_1 = 12'529^\circ$
$\bar{x}_2 = 13'450^\circ$ $\Big\} \to \bar{x}_1 - \bar{x}_2 = -0'921$

$\alpha = 2\% \Rightarrow z_{\alpha/2} = 2'325$



$$\left(-2'325 \cdot \sqrt{\frac{0'5^2}{14} + \frac{0'6^2}{6}}\ ,\ 2'325 \cdot \sqrt{\frac{0'5^2}{14} + \frac{0'6^2}{6}}\right)$$

$-0'921 \notin (-0'6487\ ,\ 0'6487)$ Valor significativo

Por lo tanto, con una significación del 2%, rechazamos la hipótesis nula. Es decir, pensamos que la graduación de los dos tipos de vino sí es distinta.