**A continuación se presentan 4 situaciones. Cada situación viene seguida por una serie de preguntas referidas a la misma así como de preguntas teóricas generales.**

**SITUACIÓN 1**: La empresa SND's de sondeos electorales ha pronosticado que el nivel de apoyo que recibirá el partido X en las próximas elecciones será del 40%. Desde el propio partido X se promueve un nuevo sondeo con el fin de contrastar la veracidad de esta afirmación. Se elige al azar una muestra aleatoria de 400 personas, con derecho a voto, de los cuales 128 manifiestan su intención de votar al partido X.

1- Con un nivel de confianza del 95%, el intervalo de confianza aproximado de la proporción de personas que votarán al partido X, es: **A) 0,2743-0,3657**; B) 0,352-0,424; C) 0,318-0,322.

*La proporción en la muestra es:* $p = \frac{128}{400} = 0{,}32$

*Al nivel de confianza del 95%, las puntuaciones típicas son:* $\pm 1{,}96$

$$p \pm Z\alpha/_2 \sqrt{\frac{p(1-p)}{n}}$$

$$0{,}32 \pm 1{,}96\sqrt{\frac{0{,}32(1-0{,}32)}{400}} \rightarrow (0{,}2743; 0{,}3657)$$

2- La hipótesis alternativa es; A) $H_1\!: \pi < 0{,}40$ ; **B)** $\boldsymbol{H_1\!: \pi \neq 0{,}40}$; C) $H_1\!: \pi > 0{,}40$.

*El contraste es bilateral porque el enunciado no especifica si el partido X espera unos resultados inferiores o superiores a los obtenidos por la empresas SND's. Simplemente, el partido X pretende contrastar la veracidad de los datos pronosticados por SND's.*

3- El estadístico de contraste es: A) -3,43; **B) -3,266**; C) -6,67.

$$Z = \frac{p - \pi_0}{\sqrt{\frac{\pi_0(1-\pi_0)}{n}}} = \frac{0{,}32 - 0{,}40}{\sqrt{\frac{0{,}40 \times 0{,}60}{400}}} = -3{,}266$$

4- Con un nivel de confianza del 99%, si los dirigentes del partido X considerasen que la proporción de apoyos no alcanza el valor pronosticado por la empresa SND's, el valor crítico para rechazar la hipótesis nula, es: **A) -2,58**; B) 1,64; C) -2,33.

*Buscando en la tabla de curva normal, en un contraste bilateral y al nivel de confianza del 99%, los valores críticos son* $\pm 2{,}58$. *Como el estadístico de contraste es negativo, el valor crítico para rechazar la hipótesis nula es -2,58.*

5- Para un contraste bilateral con un nivel de significación de 0,01, la conclusión es: **A) Rechazar la hipótesis nula con un nivel de significación del 0,01**; B) No hay evidencias para rechazar la hipótesis nula con un nivel de significación 0,05; C) No hay evidencias para rechazar la hipótesis nula con un nivel de significación 0,01.

*Rechazamos la hipótesis nula al nivel de confianza del 99% porque el estadístico de contraste es menor que el valor crítico* $(-3{,}266 < -2{,}58)$*, luego la opción A es correcta, siendo las opciones B y C claramente incorrectas.*

6- A medida que aumenta el tamaño de la muestra, la distribución muestral de la proporción se aproxima a una distribución: **A) Normal**; B) Binomial; C) $\chi^2$con $n-1$ grados de libertad.

*Pregunta teórica. Página 30 del libro de texto.*

7- La siguiente afirmación: *"La precisión del intervalo de confianza aumenta al aumentar el tamaño de la muestra"* es: A) falsa; B) verdadera para la media y la proporción y falsa para la varianza**; C) siempre verdadera**.

*A medida que aumenta el tamaño de la muestra, menor es el error típico, y cuanto menor es el error típico, mayor la precisión del intervalo de confianza. Página 45 del libro de texto.*

8- Entre dos estimadores de un mismo parámetro poblacional, es más eficiente aquel: **A) cuya distribución tenga menos variabilidad**; B) que se concentra en un rango cada vez más estrecho alrededor de su media a medida que aumenta el tamaño de la muestra; C) que utiliza toda la Información muestral relacionada con el parámetro.

*Pregunta teórica. Página 33 del libro de texto.*

**SITUACIÓN 2**. En un estudio cuyo objeto es evaluar el efecto de la música clásica sobre la capacidad de concentración disponemos de 62 sujetos con los que formamos de manera aleatoria dos grupos de 31 sujetos cada uno. Al primero de ellos (Grupo 1) se le sometió a una prueba de concentración escuchando música clásica y al segundo (Grupo 2) se le sometió a la misma prueba en condiciones normales de silencio. La puntuación media para el Grupo 1 fue de 86 puntos con una cuasivarianza igual a 150. El Grupo 2 obtuvo una media aritmética igual a 80 puntos y una cuasivarianza igual a 129. Sabiendo que la variable dependiente está medida en una escala de intervalo, asumiendo varianzas poblacionales iguales y con un nivel de confianza del 95% ¿podemos afirmar que la media en concentración es superior para el grupo que escuchó música clásica?

9- El tamaño del efecto: A) depende del tamaño muestral; **B) es independiente del tamaño muestral**; C) si es grande, necesariamente obtenemos resultados significativos.

*Página 156 del libro de texto.*

10- Si trabajamos con dos grupos siendo los tamaños de las muestras: $n_1$ = 30 y $n_2$ =35, sabemos que: A) se trata de un contraste paramétrico; B) las muestras están relacionadas; **C) las muestras son independientes**.

*La opción A es incorrecta porque el uso de un contraste paramétrico o no paramétrico no depende exclusivamente del tamaño de la muestra. Si las muestras son relacionadas, cada puntuación en una muestra tiene su pareja en la otra, por lo que en este caso $n_1 = n_2$ y por lo tanto la opción B tampoco es correcta.*

11- La hipótesis nula es: **A)** $\boldsymbol{H_0: \mu_1 - \mu_2 \leq 0}$**;** B) $H_0: \mu_1 - \mu_2 < 0$; C) $H_0: \mu_1 - \mu_2 \geq 0$.

*Queremos comprobar si la media es superior para el Grupo 1 (música clásica) que para el Grupo 2 (silencio), luego la hipótesis alternativa es* $H_1: \mu_1 - \mu_2 > 0$. *La hipótesis nula, por lo tanto, es la especificada en la opción A.*

12- El valor del estadístico de contraste es, aproximadamente: A) 1,671; **B) 2**; C) 0,508.

$$T = \frac{\bar{Y}_1 - \bar{Y}_2}{\sqrt{\frac{(n_1-1)\hat{S}_1^2 + (n_2-1)\hat{S}_2^2}{n_1+n_2-2}\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}} = \frac{86-80}{\sqrt{\frac{30\times 150 + 30\times 129}{31+31-2}\left(\frac{1}{31}+\frac{1}{31}\right)}} = 2$$

13- El tamaño del efecto es Igual a: A) 1,671; B) 2; **C) 0,508**.

$$D = \frac{\bar{Y}_1 - \bar{Y}_2}{\sqrt{\frac{(n_1-1)\hat{S}_1^2 + (n_2-1)\hat{S}_2^2}{n_1+n_2-2}}} = \frac{86-80}{\sqrt{\frac{30\times 150 + 30\times 129}{31+31-2}}} = 0{,}508$$

14- El valor crítico es igual a: **A) 1,671**; B) 2; C) 0,508.

*El contraste es unilateral derecho. Buscando en las tablas T de Student para 60 grados de libertad* $(n_1 + n_2 - 2 = 60)$ *la puntuación que supera una proporción igual a 0,95, comprobamos que la respuesta correcta es A.*

15- El nivel *p-crítico* es igual a: A) entre 0,05 y 0,10; B) entre 0,05 y 0,95; **C) 0,025**.

*Observamos en la tabla T de Student con 60 grados de libertad, que la puntuación T = 2 (estadístico de contraste) deja por debajo de sí una proporción igual a 0,975, luego el nivel p-crítico es igual a:* $1 - 0{,}975 = 0{,}025$

16- Rechazamos la hipótesis nula porque: A) el nivel de significación es menor que el nivel *p-crítico;* ***B) e*l estadístico de contraste es mayor que el valor crítico**; C) no rechazamos la hipótesis nula.

*Claramente la opción correcta es B. La opción A es falsa porque si el nivel de significación es menor que el nivel crítico mantenemos la hipótesis nula.*

**SITUACIÓN 3**. Un agrónomo desea conocer el efecto que sobre el rendimiento de una variedad de trigo tiene la adición de 3 tipos diferentes de fosfatos al terreno. Para ello parcela un terreno en 12 áreas del mismo tamaño y trata cada cuatro parcelas con un tipo distinto de fosfato (A, B o C). A continuación siembra trigo en cada uno de ellos y, después de la recolección, mide la cantidad de trigo producida por cada superficie (en hectólitros por hectárea o hl/ha). La Tabla 1 muestra el rendimiento obtenido en función del tipo de fosfato añadido.

| | | | | |
|---|---|---|---|---|
| Fosfato A | 48 | 49 | 50 | 49 |
| Fosfato B | 47 | 49 | 48 | 48 |
| Fosfato C | 49 | 51 | 50 | 50 |

Sabemos que: $SC_{S/A} = SS_{dentro\_niveles} = 6$; $SC_A = SS_{entre\_niveles} = 8$

*Ordenando los datos que tenemos en una tabla:*

| | SC | gl | MC | F |
|---|---|---|---|---|
| Entre niveles | 8 | I-1=2 | 8/2=4 | $\frac{4}{0{,}67} = 6$ |
| Dentro niveles | 6 | N-I=9 | 6/9=0,67 | |
| Total | 8+6=14 | N-1=11 | | |

17- La hipótesis nula es A) $H_0: \mu_{fosfatoA} = \mu_{fosfatoB}$; **B)** $\boldsymbol{H_0: \mu_{fosfatoA} = \mu_{fosfatoB} = \mu_{fosfatoC}}$**;** C) $H_0: \mu_{fosfatoA} \neq \mu_{fosfatoB} \neq \mu_{fosfatoC}$

*La hipótesis nula ha de postular la igualdad de las medias en todos los niveles, luego la opción correcta es B.*

18- Los grados de libertad del estadístico F de contraste valen: A) (3,9); B) (2,11); **C) (2,9)**.

*Como comprobamos en la tabla de ANOVA, la opción correcta es C.*

19- El estadístico de contraste para evaluar la significatividad del "Tipo de Fosfato" vale aproximadamente: **A) 6**; B) 3.295; C) 0.392.

*Como comprobamos en la tabla de ANOVA, la opción correcta es A.*

20- Se trata de un diseño: A) de una muestra con 12 observaciones; **B) de un factor con tres niveles (tipo de fosfato)**; C) de un factor (parcela de terreno) con cuatro niveles.

*Claramente la opción correcta es B.*

**SITUACIÓN 4**. En un colegio se estudia la relación que puede existir entre la calificación global de los estudiantes al final de curso (Y) y la obtenida el curso anterior (X). Efectuado un análisis de regresión, con una muestra de estudiantes, la prueba para determinar la significación del modelo de regresión ajustado es la siguiente:

| | SC | gl | MC | F | Prob. |
|---|---|---|---|---|---|
| Regresión | 42,734 | 1 | 42,734 | 17,345 | # |
| Residual | U | V | W | | |
| Total | 64,909 | 10 | | | |

*Deducimos fácilmente los datos que nos faltan en la tabla:*

$U = 64{,}909 - 42{,}734 = 22{,}175;\ V = 10 - 1 = 9;\ \ W = {}^{U}\!/_{V} = {}^{22{,}175}\!/_{9} = 2{,}4639$

*La tabla completa queda:*

| | SC | gl | MC | F |
|---|---|---|---|---|
| Regresión | 42,734 | 1 | 42,734 | 17,345 |
| Residual | 22,175 | 9 | 2,4639 | |
| Total | 64,909 | 10 | | |

21- Uno de los supuestos básicos del análisis de regresión simple es: **A) los pronósticos y los errores son independientes**; B) las distribuciones condicionadas de los errores deben de tener una distribución uniforme; C) no deben estar relacionadas las variables predictora y la variable dependiente o criterio.

*Página 343 del libro de texto.*

22- ¿Aproximadamente cuál es el coeficiente de correlación de Pearson entre X e Y? A) 0,787; B) 0,97; **C) 0,811**.

*El cociente entre la suma de cuadrados de la regresión y la suma de cuadrados total es el coeficiente de determinación, luego su raíz cuadrada será el coeficiente de correlación de Pearson.*

$$r_{xy} = \sqrt{\frac{42{,}734}{64{,}909}} = 0{,}811$$

23- ¿Aproximadamente, cuál es la varianza muestral insesgada de los errores del modelo de regresión? A) 2,217; B) 4,273; **C) 2,464**.

*La varianza muestral insesgada de los errores del modelo de regresión coincide con la media cuadrática residual.*

24- ¿Con un nivel de confianza del 95%, cuál es el valor de la F teórica, por encima del cual se debería rechazar la hipótesis de que hay no relación lineal significativa entre Y y X?; **A) 5,117**; B) 5,59; C) 6,608.

*Buscando en las tablas F de Fisher para 1 y 9 grados de libertad, comprobamos que la respuesta correcta es A.*

25- ¿Cuál es el error típico del modelo de regresión? A) 2,464; **B) 1,57**; C) 1,665.

$$\sigma_e = \sqrt{\frac{\sum(Y - Y')^2}{n - p - 1}} = \sqrt{\frac{22{,}175}{9}} = 1{,}57$$