Cálculo de probabilidades y estadística **Examen Final Mayo**

EJERCICIO 1 (SOLUCIONES) **5 de junio de 2025**

1.– Se modela un sistema ecológico de manera que la especie animal dominante consume untanto por uno, X, $0 \leq X \leq 1$, de los recursos totales del sistema que denotamos R. Se ha supuesto que X tiene la función de densidad de probabilidad:

$$f_X(x) = 2(1-x), \quad 0 \leq x \leq 1$$

a) Hallar la función de densidad de probabilidad de W, el total de recursos consumido por la especie animal.

b) Suponiendo que se dispone de una muestra de W, $w_1, w_2, ...w_n$, determinar cómo se calcularía el estimador de máxima verosimilitud de R. Particularizar el resultado para el caso de una muestra de tamaño 1.

SOLUCIÓN

a) El total de los recursos consumidos por la especie animal es $W = XR$. Como esta relación es creciente en X,

$$f_W(w) = \left|\frac{dx}{dw}\right| f_X\left(\frac{w}{R}\right) = \frac{1}{R}2(1-\frac{w}{R}) = \frac{2(R-w)}{R^2}, \quad 0 \leq w \leq R$$

y es obvio que

$$\int_0^R \frac{2(R-w)}{R^2} \; dw = \frac{2}{R^2}\left[Rw - \frac{w^2}{2}\right]_0^R = 1$$

b) La verosimilitud de una muestra de W, como función del parámetro R es

$$L(R|w_1, w_2, ...w_n) = \prod_{i=1}^{n} \frac{2(R-w_i)}{R^2}$$

Tomando logaritmos

$$\ln L(R|w_1, w_2, ...w_n) = \sum_{i=1}^{n} \ln \frac{2(R-w_i)}{R^2} = \sum_{i=1}^{n} \left[\ln(2(R-w_i)) - 2\ln R\right]$$

Derivando con respecto a R

$$\frac{d\ln L}{dR} = \sum_{i=1}^{n}\left(\frac{1}{R-w_i}\right) - \frac{2n}{R}$$

y la ecuación a resolver para calcular el estimador de máxima verosimilitud de R sería

$$\sum_{i=1}^{n}\left(\frac{1}{\hat{R}-w_i}\right) - \frac{2n}{\hat{R}} = 0$$

En general, esta ecuación se resolvería para un conjunto de valores $w_1, w_2, ...w_n$ mediante cualquier método numérico. Para el caso $n = 1$

$$\frac{1}{\hat{R}-w_1} - \frac{2}{\hat{R}} = 0 \quad \Longrightarrow \quad \hat{R} = 2w_1$$

2.– Para calcular la profundidad de la roca que se encuentra bajo un terreno arcilloso van a utilizarse ondas de choque producidas por geófonos. Estos aparatos miden la frecuencia de recepción de una onda emitida, deduciendo la profundidad de la roca a partir de esa frecuencia. Para el estudio van a usarse dos tipos de geófonos. Con el primer geófono se han hecho $n_1 = 20$ medidas y la frecuencia media obtenida ha sido $\bar{x}_1 = 3.4 \times 10^3\ Hz$ y su desviación típica $S_1 = 0.2 \times 10^3\ Hz$. Con el segundo geófono se han hecho $n_2 = 15$ medidas obteniéndose $\bar{x}_2 = 3.1 \times 10^3\ Hz$ y $S_2 = 0.15 \times 10^3\ Hz$. Para calibrar los aparatos interesa hallar un intervalo de confianza $(1-\alpha = 0.99)$ sobre la diferencia de medias de las frecuencias de ambos geófonos.

Calcular dicho intervalo suponiendo que las desviaciones típicas son las verdaderas $(S_1 = \sigma_1,\quad S_2 = \sigma_2)$.

——————————SOLUCIÓN——————————

a) Si las desviaciones típicas son las verdaderas utilizaremos como estadístico

$$\frac{\bar{x}_1 - \bar{x}_2 - (m_1 - m_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \equiv N(0,1)$$

Entonces, para $\alpha = 1\,\%$, el intervalo de confianza en la diferencias de medias es

$$IC_{99\,\%} = \left[(\bar{x}_1 - \bar{x}_2) - 2.576\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}, (\bar{x}_1 - \bar{x}_2) + 2.576\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}\right]$$

Sustituyendo valores

$$\boxed{IC_{99\,\%} = [0.1476 \times 10^3, 0.4524 \times 10^3]}$$

3.– Un fabricante de bombillas asegura que su modelo "Moltallum ML-500W" tiene una vida media de 1600 horas. Un comerciante toma una muestra de 100 de esas bombillas y encuentra que la vida media muestral es de 1570 horas con una desviación típica muestral de 120 horas.

a) Contrastar la aseveración del fabricante contra la hipótesis alternativa de *media diferente de 1600 horas*, usando un nivel de significación de 0.01.

b) Idem contra la alternativa *media menor que 1600 horas* usando un nivel de significación de 0.01.

c) ¿Cuál es el nivel p de los contrastes indicados en los dos apartados anteriores?

d) ¿Cuál es la probabilidad de error tipo II en cada caso, si la hipótesis alternativa se concreta en *media igual a 1570 horas?*

e) ¿Qué contraste de hipótesis debería usar el comerciante comprador? Indicar y razonar únicamente, sin operaciones ni cálculos.

——————SOLUCIÓN——————

a) El contraste de hipótesis que se plantea es

$H_o : m = m_o = 1600$

$H_1 : m \neq m_o = 1600$

Como se supone σ desconocido, y sabiendo que la región crítica es simétrica dada la hipótesis alternativa indicada, podemos escribir

$$1 - \alpha = P\Big[- a < \frac{\overline{x} - m_o}{S/\sqrt{n-1}} \leq a \Big| H_o \Big] = F_T(a) - F_T(-a) = 2F_T(a) - 1$$

ya que en H_o el estadístico $\frac{\overline{x}-m_o}{S/\sqrt{n-1}}$ es una variable T_{n-1}. De esta expresión obtenemos $a = 2.6264$ y como nuestro estadístico vale $\frac{\overline{x}-m_o}{S/\sqrt{n-1}} = -2.4875$ está en la región de aceptación ($R_A = [-2.6264, 2.6264]$) podemos aceptar la hipótesis primaria planteada por el fabricante.

b) En este caso, la hipótesis alternativa cambia y el contraste es del tipo

$H_o : m = m_o = 1600$

$H_1 : m < m_o = 1600$

Aquí es evidente que la región crítica está a la izquierda de la distribución. Por tanto

$$1 - \alpha = P\Big[- a < \frac{\overline{x} - m_o}{S/\sqrt{n-1}} \Big| H_o \Big] = 1 - F_T(-a) = F_T(a)$$

y obtenemos $a = 2.3646$. Luego no podemos aceptar la hipótesis primaria ya que nuestro estadístico está fuera de la región de aceptación ($R_A = [-2.3646, \infty]$)

c) El nivel p (máximo nivel de significación a que podemos aceptar la hipótesis primaria dada la muestra (estadístico de control 2.4875)) es de $\boxed{p = 0.0145 = 1.45\,\%}$ en el primer caso y $\boxed{p = 0.00725 = 0.725\,\%}$ en el segundo, bajísimos ambos.

d) En el primer caso, la probabilidad del error tipo II es

$$\beta = P\Big[- a < \frac{\overline{x} - m_o}{S/\sqrt{n-1}} \leq a \Big| H_1 \Big] = P\Big[- a + \frac{m_o - m_1}{S/\sqrt{n-1}} < \frac{\overline{x} - m_1}{S/\sqrt{n-1}} \leq a + \frac{m_o - m_1}{S/\sqrt{n-1}} \Big| H_1 \Big] =$$

$$= F_T\Big(a + \frac{m_o - m_1}{S/\sqrt{n-1}}\Big) - F_T\Big(-a + \frac{m_o - m_1}{S/\sqrt{n-1}}\Big)$$

Sustituyendo valores, con $m_1 = 1570$ obtenemos

$$\boxed{\beta = F_T(5.1139) - F_T(-0.1389) \approx 1 - 0.4449 = 0.5551}$$

En el segundo caso tenemos

$$\beta = P\Big[-a < \frac{\overline{x} - m_o}{S/\sqrt{n-1}}\Big|H_1\Big] = P\Big[-a + \frac{m_o - m_1}{S/\sqrt{n-1}} < \frac{\overline{x} - m_1}{S/\sqrt{n-1}}\Big|H_1\Big] = 1 - F_T\Big(-a + \frac{m_o - m_1}{S/\sqrt{n-1}}\Big)$$

y operando

$$\boxed{\beta = 1 - F_T(0.1229) = 0.5488}$$

e) Si el comprador sospecha alguna falsedad en la descripción de la mercancía debería plantear el siguiente contraste:

$H_o : m \leq m_o = 1600$ (las bombillas son peores de lo que se dice)

$H_1 : m > m_o = 1600$ (las bombillas son mejores de lo que se dice)

Es obvio que el nivel p de este contraste con los datos obtenidos es altísimo (de hecho, $p \approx 99\,\%$) lo que indica que el fabricante miente y que esas bombillas (a pesar del rimbombante nombre) no se deben comprar