Cálculo de probabilidades y estadística **Examen Segundo Parcial**

EJERCICIO 1 (SOLUCIONES) **21 de mayo de 2025**

1.– La nota de Cálculo de Probabilidades y Estadística de un estudiante elegido al azar tiene por función de densidad

$$f_X(x) = \frac{3x^2}{\theta^3}, \quad 0 \leq x \leq \theta, \quad \theta > 0$$

En el grupo hay un total de n=25 estudiantes. Solamente se conoce la nota máxima de la clase. Se pide:

a) Determinar la distribución de la nota máxima de la clase.

b) Determinar un estimador de máxima verosimilitud para θ. (Se recuerda que sólo se conoce la máxima nota de la clase)

c) Determinar la cota superior de θ con una confianza del 90 % si se sabe que la nota máxima de la clase es 9.25.

——————————SOLUCIÓN——————————

a) Si llamamos $Y = \max\limits_{i=1}^{n} (X_i)$ sabemos que la función de distribución acumulada de Y vale $F_Y(y) = \left[F_X(y)\right]^n$ (siempre que supongamos las variables X_i (nota de cada estudiante) independientes e igualmente distribuidas). Pero como $F_X(x) = \frac{x^3}{\theta^3}$ (se calcula sin más que integrar $f_X(x)$ entre 0 y x), obtenemos

$$\boxed{F_Y(y) = \frac{y^{3n}}{\theta^{3n}}, \quad 0 \leq y \leq \theta, \quad \theta > 0}$$

y sin más que derivar

$$\boxed{f_Y(y) = \frac{3ny^{3n-1}}{\theta^{3n}}, \quad 0 \leq y \leq \theta, \quad \theta > 0}$$

b) La función de verosimilitud, para una muestra de tamaño 1 (obsérvese que nuestra variable aleatoria es ahora el máximo de las n notas de la clase) es

$$L = \frac{3ny_1^{3n-1}}{\theta^{3n}}$$

Esta función se hace tanto mas grande cuanto menor sea θ, luego el parámetro θ ha de ser lo más pequeño posible. Pero por otra parte, para que la función de verosimilitud sea diferente de cero, ha de verificarse $y \leq \theta$. Luego el menor valor de θ que no anula la función de verosimilitud es

$$\boxed{\hat{\theta} = y_1 = \max\limits_{i=1}^{n} (x_i)}$$

c)Calcular la cota superior de θ con una confianza del $1-\alpha$ es lo mismo que calcular un intervalo de confianza por un sólo lado sobre θ

Para poner calcular un intervalo de confianza sobre θ necesitamos encontrar una variable relacionada con $\hat{\theta}$ pero cuya distribución no dependa de θ. Para simplificar la notación vamos a llamar $T = \hat{\theta}$. Sabemos que como $T = y_1$, la distribución de T será la misma que de Y. Es decir

$$f_T(t) = \frac{3nt^{3n-1}}{\theta^{3n}}, \quad 0 \leq t \leq \theta$$

Es fácil comprobar que $E[T] = A_1\theta$ y $Var[T] = A_2\theta^2$ donde A_1 y A_2 son constantes. Por tanto, si elegimos un cambio de variable de la forma $Z = \frac{T}{\theta}$, por lo pronto $E[Z]$ y $Var[Z]$ no dependerán de θ. Veamos cual es la distribución de esa variable Z

$$f_Z(z) = \left|\frac{dt}{dz}\right| f_T(t) = \theta\frac{3nt^{3n-1}}{\theta^{3n}} = \frac{3n(z\theta)^{3n-1}}{\theta^{3n-1}} = 3nz^{3n-1}, \quad 0 \leq z \leq 1$$

e integrando

$$F_Z(z) = z^{3n}, \quad 0 \leq z \leq 1$$

Efectivamente, esta variable no depende del parámetro, luego podemos calcular probabilidades con ella. Para hallar un intervalo de confianza sobre θ actuaremos de la siguiente forma

$$1-\alpha = 0.9 = P[Z > a] = P[\frac{T}{\theta} > a] = P[\theta \leq \frac{T}{a}] = P[\theta \leq \frac{\hat{\theta}}{a}]$$

Pero

$$P[Z > a] = 1 - F_Z(a) = 1 - a^{3n} \quad \Rightarrow \quad a^{3n} = \alpha \quad \Rightarrow \quad a = \alpha^{\frac{1}{3n}}$$

Y para $T = \hat{\theta} = 9.25$, $1-\alpha = 0.9$ y $n = 25$ obtenemos $a = 0.9698$ y por lo tanto el intervalo de confianza (cota superior de θ) es

$$\boxed{\theta \leq \frac{9.25}{0.9698} = 9.5388}$$

2.– Una fábrica vierte a una ría a través de un colector las aguas procedentes de un proceso industrial contaminante tras aplicar los oportunos tratamientos de depuración. Las agencias de salud vigilan de forma periódica la calidad del agua vertida mediante la toma y el análisis de muestras. Según la legislación vigente se considera que la cantidad promedio de amoniaco en el agua vertida no debe exceder el valor de 2 ppm (partes por millón o mg/L), con una seguridad del 99 %.

Se pide establecer el contraste a efectuar para comprobar la calidad del agua vertida. Coméntese el significado físico de cada error. Para comprobar la calidad del agua vertida, se toma una muestra de tamaño 16, la cual proporciona una media y una desviación típica muestral insesgada de 1.5 y 1.2 ppm, respectivamente. ¿Se puede considerar que la fábrica cumple la legislación vigente en lo referente a la calidad del agua vertida? Calcúlese el nivel p obtenido.

————————SOLUCIÓN————————

a) Si establecemos el contraste de hipótesis

$H_0 : m \leq 2$

$H_1 : m > 2$

con m en ppm, los errores α y β tendrán el siguiente significado:

$\alpha : P[\text{rechazar } H_0 | H_0 \text{ es cierta}] = P[\text{rechazar el agua siendo ok}]$

$\beta : P[\text{aceptar } H_0 | H_1 \text{ es cierta}] = P[\text{aceptar el agua no siendo ok}]$

Luego, en este caso, el riesgo potencial está en el error $TipoII$.

Si en cambio, establecemos el contraste de hipótesis como

$H_0 : m > 2$

$H_1 : m \leq 2$

los errores α y β tendrán el siguiente significado:

$\alpha : P[\text{rechazar } H_0 | H_0 \text{ es cierta}] = P[\text{aceptar el agua siendo no aceptable}]$

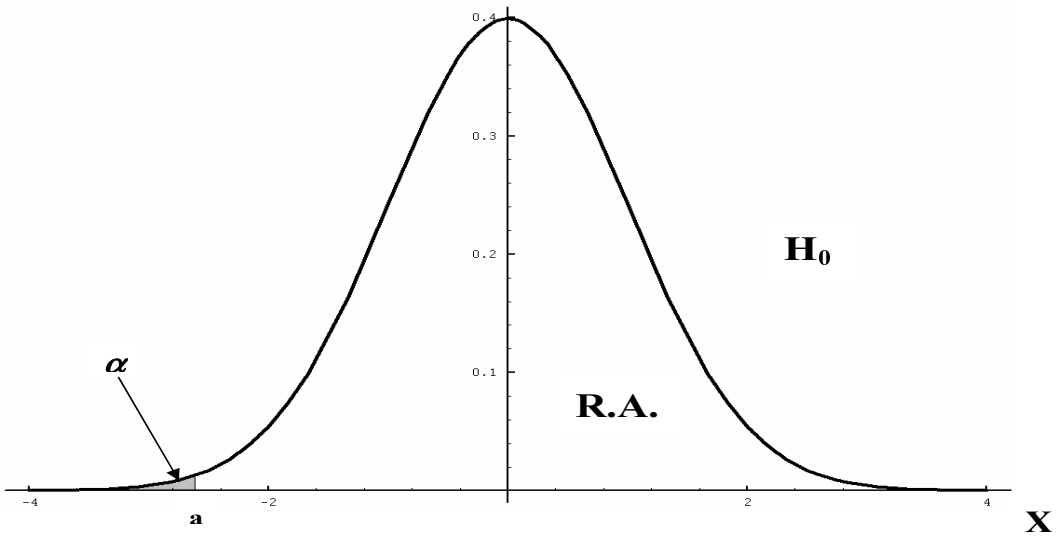
$\beta : P[\text{aceptar } H_0 | H_1 \text{ es cierta}] = P[\text{rechazar el agua siendo ok}]$

En este caso, la decisión es ajustada a α. Por lo tanto, utilizaremos el segundo test de hipótesis.

b) Suponiendo la distribución normal, y dado que desconocemos la varianza de la población, utilizaremos el estadístico

$$T = \frac{\bar{x} - m_x}{S^*/\sqrt{n}} \,|H_0 \quad \Rightarrow \quad t_{n-1}$$

siendo la región crítica la especificada en la siguiente figura:



Para $\alpha = 0.01$, se obtiene $a = -2.6$

Para rechazar H_0 hace falta que $\bar{x} < m_x + a\ \frac{S^*}{\sqrt{n}} = 2 - 2.6\ \frac{1.2}{\sqrt{16}} = 1.22$

Como $m = 1.5 > 1.22$ no podemos rechazar que el agua esté contaminada. El nivel p obtenido es

$$p = P\left[T \leq \frac{1.5 - 2}{1.2/\sqrt{16}}\ \middle|H_0\right] = P[T \leq -1.67|H_0]$$

que para 15 grados de libertad resulta $p = 1 - 0.94 = 6\,\%$

3.– La duración en horas, X, de unas lámparas eléctricas fabricadas por una cierta empresa, puede considerarse distribuida normalmente $N(m_x, \sigma_x^2)$ con $\sigma_x = 250$ horas. Para realizar el control de calidad se toman 10 lámparas de una misma partida, obteniendo las duraciones $x_1, x_2, \ldots, x_{10}$. Se pretende contrastar la hipótesis $H_o: m_x \geq 2000$ frente a $H_1: m_x < 2000$ haciendo uso del estimador de máxima verosimilitud

$$\hat{m} = \bar{x} = \frac{1}{10}\sum_{i=1}^{10} x_i$$

a) Calcular el nivel de significación α del contraste si la región de aceptación es $\bar{x} \geq 1870$.

b) Para evitar cálculos adicionales y acelerar el proceso de control, se quiere sustituir este contraste por uno basado en el estadístico

$$M = min[x_1, x_2, \ldots, x_{20}]$$

que utiliza una muestra con el doble de lámparas. Calcular la nueva región de aceptación para el mismo nivel de significación del apartado anterior.

c) Calcular la potencia de ambos contrastes cuando la verdadera media es $m_x = 1800$ horas, y compararlos.

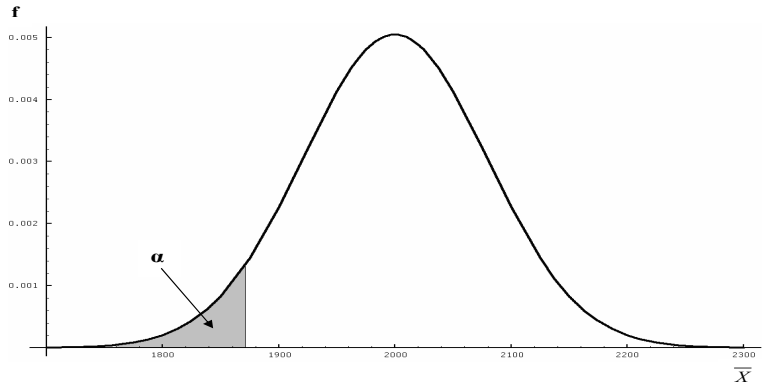
——————————SOLUCIÓN——————————

a) El contraste de hipótesis que nos piden es

$H_0 : m_x \geq 2000$

$H_1 : m_x < 2000$

La media muestral, $\bar{x}$ tiene distribución $N(m_x, \sigma_x/\sqrt{n})$. La región crítica se encuentra obviamente a la izquierda de la distribución, como puede verse en la figura siguiente particularizada para H_0.



Por lo tanto, el valor de α se calcularía como

$$\alpha = F_{\bar{X}}(1870) = F_U\Big(\frac{1870 - 2000}{250/\sqrt{10}}\Big) = F_U(-1.6444) = 0.05$$

Luego

$$\boxed{\alpha = 0.05}$$

b) Ahora el estadístico a usar es $M = min[x_1, x_2, \ldots, x_{20}]$. La distribución de este estadístico se calcula de la siguiente manera (siempre que supongamos que la muestra $x_1 ... x_{20}$ es aleatoria):

$$P[M > m] = P[x_1 > m \cap x_2 > m \cap x_3 > m \cap ... \cap x_{20} > m] = [1 - F_X(m)]^{20}$$

Luego

$$F_M(m) = 1 - P[M > m] = 1 - [1 - F_X(m)]^{20}, \quad m \in \mathbb{R}$$

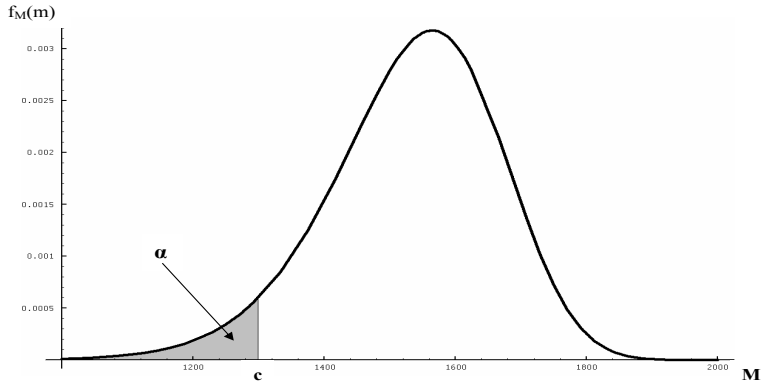
Dado que el contraste de hipótesis es el mismo, la región crítica estará también situada en la cola inferior de la distribución. Luego para evaluar el valor límite de la región crítica , c, debemos calcular

$$\alpha = 0.05 = F_M(c) = 1 - [1 - F_X(c)]^{20} \quad \Longrightarrow \quad F_X(c) = 1 - (1 - \alpha)^{\frac{1}{20}} = 0.002561$$

Pero, en H_0, $X \equiv N(m_x, \sigma_x) = N(2000, \sigma_x = 250)$. Por tanto

$$0.002561 = F_X(c) = F_U\Big(\frac{c - 2000}{250}\Big) \quad \Longrightarrow \quad \frac{c - 2000}{250} = -2.7994 \quad \Longrightarrow \quad \boxed{c = 1300.15}$$

Y nuestra regla de actuación para este segundo estadístico será rechazar H_0 si $M < 1300.15$. En la siguiente figura se muestra la región crítica para este estadístico



c) La potencia para el primer estadístico es

$$\Pi = 1 - \beta = 1 - P[\bar{x} \in R_A | H_1] = 1 - P[\bar{x} \geq 1870 | H_1] = P[\bar{x} \leq 1870 | H_1] =$$

$$P\Big[\frac{\bar{x} - 1800}{250/\sqrt{10}} < \frac{1870 - 1800}{250/\sqrt{10}}\Big] = F_U\Big[\frac{1870 - 1800}{250/\sqrt{10}}\Big] = F_U[0.88544] = \boxed{0.812 = \Pi}$$

La potencia para el segundo estadístico será

$$\Pi = 1 - \beta = 1 - P[M \in R_A | H_1] = 1 - P[M \geq 1300.15 | H_1] =$$

$$= P[M \leq 1300.15 | H_1] = F_{M|H_1}(1300.15) = 1 - [1 - F_{X|H_1}(1300.15)]^{20}$$

Pero

$$F_{X|H_1}(1300.15) = F_U\Big[\frac{1300.15 - 1800}{250}\Big] = F_U[-1.9994] = 0.0228$$

Y por lo tanto

$$\boxed{\Pi = 1 - [1 - 0.0228]^{20} = 0.3695}$$

Es decir que duplicar el tamaño de la muestra no es suficiente para igualar la potencia del contraste definido por el estimador de máxima verosimilitud (a igualdad de nivel de confianza).