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

EJERCICIO 1 (SOLUCIONES) **20 de mayo de 2024**

1.– Un conductor tarda un tiempo X en pagar un peaje. Se supone que X responde a una distribución exponencial de parámetro λ trasladada al rango $[3,\infty)$ (en segundos). Para estimar el parámetro λ se espera que paguen 100 conductores y se anota el tiempo que ha tardado el conductor más rápido. Esta operación se repite n veces para obtener los tiempos mínimos $y_1, y_2, ..., y_n$. Se pide:

a) Hallar la distribución de Y, tiempo mínimo de pago entre 100 conductores.

b) Establecer la ecuación que debe satisfacer el estimador de máxima verosimilitud de λ, basado en los datos $y_1, y_2, ..., y_n$.

c) Dar un intervalo de confianza del 90 % sobre λ, de la forma $IC_{1-\alpha} = (\lambda \geq a)$ en el caso $n = 1$, $Y_1 = 3.2$.

d) ¿Qué debe hacerse en el apartado c) si $n = 500$?

———————SOLUCIÓN———————

Sea $x_1,\ x_2,\ ...,\ x_{100}$ el tiempo de pago de los 100 conductores. Sabemos que

$$f_X(x) = \lambda e^{-\lambda(x-3)}, \quad F_X(x) = 1 - e^{-\lambda(x-3)}, \quad x \geq 3$$

a) Sea Y el tiempo mínimo de pago entre 100 conductores. Calculemos la distribución de Y suponiendo a los 100 conductores independientes

$$P[Y \geq y] = \big[P[X \geq y]\big]^{100} = [1 - F_X(y)]^{100} = [1 - (1 - e^{-\lambda(y-3)})]^{100} = e^{-100\lambda(y-3)} = 1 - F_Y(y)$$

por tanto

$$\boxed{F_Y(y) = 1 - e^{-100\lambda(y-3)}, \quad y \geq 3}$$

b) Si tenemos una muestra de tamaño n la función de verosimilitud es

$$L(y_1, y_2, ..., y_n|\lambda) = \prod_{i=1}^{n} f_{Y_i}(y_i) = \prod_{i=1}^{n} 100\lambda e^{-100\lambda(y_i-3)} = 100^n \lambda^n e^{-100\lambda \sum(y_i-3)}$$

Tomando logaritmos

$$\ln L(y_1, y_2, ..., y_n|\lambda) = n \ln 100 + n \ln \lambda - 100\lambda \sum(y_i - 3)$$

Derivando e igualando a cero,

$$\frac{\partial \ln L}{\partial \lambda} = \frac{n}{\hat{\lambda}} - 100 \sum(y_i - 3) = 0 \quad \Rightarrow \quad \boxed{\hat{\lambda} = \frac{n}{100 \sum(y_i - 3)}}$$

c) Si $n = 1$, sea $y_1 = 3.2$ el valor de la muestra. Entonces el estimador es

$$\hat{\lambda} = \frac{1}{100(y_1 - 3)}$$

Si calculamos un intervalo de confianza sobre un solo lado (que es el que nos interesa en este caso), buscamos el valor de b sobre el que podamos afirmar que $\hat{\lambda} > b$, con una confianza del 90 %. Obsérvese que esto equivale a dar una cota máxima sobre el tiempo medio de pago. Sabemos que la distribución de y_1 es

$$f_Y(y_1) = 100\lambda e^{-100\lambda(y_1 - 3)}, \quad y_1 \geq 3$$

Definamos la variable $U = 100\lambda(y - 3)$ Veamos la distribución de esta variable. Obviamente $U \geq 0$ y

$$f_U(u) = \frac{dY}{dU} f_Y(y) = \frac{1}{100\lambda} 100\lambda e^{-100\lambda(y_1 - 3)} = e^{-u}$$

luego $U \equiv EX[1]$. Para calcular ahora el intervalo de confianza sobre λ hacemos como siempre:

$$0.9 = P[U > k] = P[100\lambda(y_1 - 3) > k] = P\Big[\lambda > \frac{k}{100(y_1 - 3)}\Big]$$

Pero

$$0.9 = P[U > k] = 1 - F_U(k) = 1 - (1 - e^{-k}) = e^{-k} \quad \Longrightarrow \quad k = 0.10536$$

y el intervalo de confianza pedido resulta ser

$$\boxed{IC_{90\,\%} \ \Rightarrow \ \lambda > 0.005268}$$

d) Si hubiéramos tenido una muestra de tamaño 500 hubiéramos podido suponer que la distribución del estadístico

$$W = \frac{\sum_{i=1}^{500} (y_i - 3)}{500} \equiv \text{Normal}$$

Dado que $z_i = (y_i - 3)$ es una exponencial con parámetros 100λ, la distribución de W es

$$W \equiv N\Big(\frac{1}{100\lambda}, \frac{1}{500(100\lambda)^2}\Big)$$

y a partir de aquí se calcularía todo como siempre.

2.– Se están analizando dos materiales de aleación de aluminio como refuerzo para las alas de un avión ligero de transporte. Los resultados de ensayos de rotura a tracción de elementos fabricados con estos dos materiales han sido los siguientes, en Mpa:

Tipo de material	Tamaño de la muestra	Media muestral	Desviación típica poblacional
1	10	876	10
2	12	745	15

Con estos resultados, calcúlese intervalos de confianza del 90 %, 95 % y 99 % sobre la diferencia de medias de las dos poblaciones, $m_1 - m_2$. A la vista de lo obtenido, ¿qué puede decirse de los dos materiales desde el punto de vista de la resistencia a tracción? Puede considerarse que la resistencia a tracción de ambos materiales tiene una distribución normal.

————————SOLUCIÓN————————

Sabemos que la distribución de la media muestral, si la población subyacente es normal (o la muestra lo suficientemente grande) es $N\big(m, \frac{\sigma^2}{n}\big)$. En nuestro caso tenemos dos muestras de diferentes materiales que proporcionan las medias muestrales $\bar{x}_1$ y $\bar{x}_2$. Entonces, la distribución de su diferencia es

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

si suponemos la muestras independientes (lo que es lógico). Por lo tanto, el intervalo de confianza sobre las diferencias de las medias es

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

Los valores de c se obtienen de las tablas de la distribución normal, con

$$c = 1.645 \text{ para } 1-\alpha = 0.9, \quad c = 1.96 \text{ para } 1-\alpha = 0.95, \quad c = 2.576 \text{ para } 1-\alpha = 0.99$$

Operando en la expresión anterior obtenemos

$$\boxed{IC_{90\,\%} = [122.2, 139.8] \qquad IC_{95\,\%} = [120.5, 141.5] \qquad IC_{99\,\%} = [117.2, 144.8]}$$

Con lo que podemos asegurar que el material 1 tiene una resistencia a tracción superior en por lo menos 117 Mpa al material 2. Obsérvese que, adicionalmente, el material 1 es más homogéneo. Luego es claramente mejor.

3.– En ocasiones los productos radiactivos de desecho industrial van a dar a las fuentes de agua que se utilizan para el consumo de la población. Por razones como ésta, las agencias de salud vigilan de forma periódica las fuentes naturales de agua mediante la toma y el análisis de muestras de agua. Legalmente se considera que la cantidad promedio de radiación en el agua para beber no debe exceder el valor de 4 picocuries por litro de agua, con una seguridad del 99 %.

Se pide establecer el contraste a efectuar para comprobar la potabilidad de las aguas. Coméntese el significado físico de cada error. Para comprobar la potabilidad de una fuente, se toma una muestra de tamaño 16, la cual proporciona una media y una desviación típica muestral insesgada de 3.5 y 1.2 picocuries, respectivamente. ¿Se puede considerar potable el agua analizada? Calcúlese el nivel p obtenido.

Nota: Supóngase que la cantidad de radiación por litro de agua se comporta, de forma aproximada, como una variable normal.

——————————SOLUCIÓN——————————

a) Si establecemos el contraste de hipótesis

$H_0 : m \leq 4$

$H_1 : m > 4$

con m en picocuries por litro de agua, los errores α y β tendrán el siguiente significado:

α : P[rechazar $H_0|H_0$ es cierta] = P[rechazar el agua siendo potable]

β : P[aceptar $H_0|H_1$ es cierta] = P[aceptar el agua no siendo potable]

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

Si en cambio, establecemos el contraste de hipótesis como

$H_0 : m > 4$

$H_1 : m \leq 4$

los errores α y β tendrán el siguiente significado:

α : P[rechazar $H_0|H_0$ es cierta] = P[aceptar el agua siendo no potable]

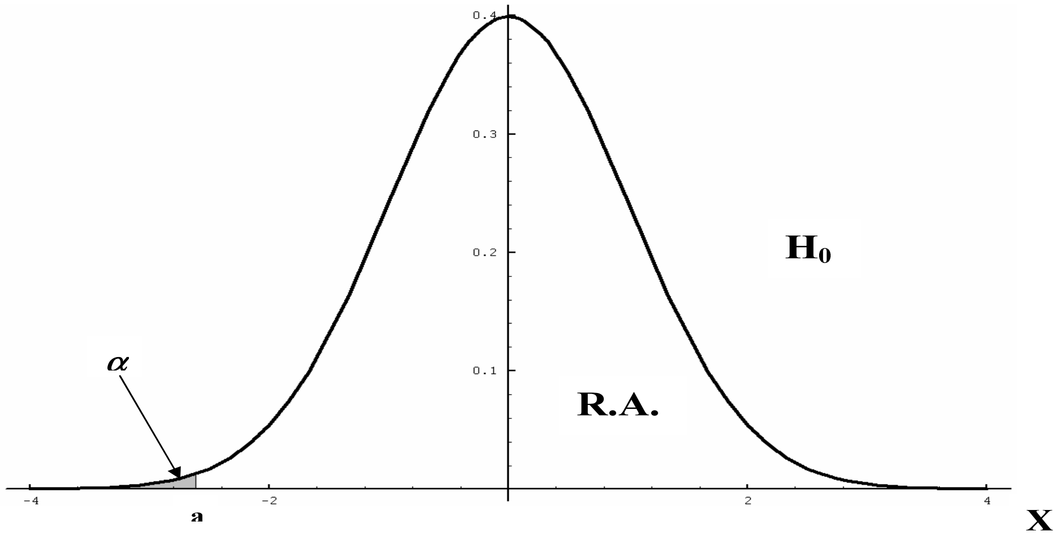
β : P[aceptar $H_0|H_1$ es cierta] = P[rechazar el agua siendo potable]

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

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

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 $\bar{x} < m_x + a\ \frac{S^*}{\sqrt{n}} = 4 - 2.6\ \frac{1.2}{\sqrt{16}} = 3.22$

Como $m = 3.5 > 3.22$ no podemos rechazar que el agua esté contaminada. El nivel p obtenido es

$$p = P\left[T \leq \frac{3.5 - 4}{1.2/\sqrt{16}}\ \big|H_0\right] = P[T \leq -1.67|H_0]$$

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