CPE (SEGUNDO CURSO)

PRÁCTICA 15 **SOLUCIONES** (**Curso 2023–2024**)

1.– Los valores observados en una muestra de extensión 14 para la tensión de rotura de cables de kevlar utilizados como tirantes de estructuras metálicas son, en GPa,

9.9	6	5.2	7.3	11.8	10.3	8.2
7.5	6.6	12.6	16.8	12.3	9.8	10.3

Determinar intervalos de confianza para la tensión media de rotura, así como para la varianza poblacional de la variable tensión de rotura.

———————SOLUCIÓN———————

A partir de los datos de la muestra obtenemos

$$n = 14, \quad \overline{X} = 9.61429, \quad S_X = 3.01256 \quad S_X^2 = 9.07551$$

Suponemos que la variable "tensión de rotura" se distribuye normalmente.

a) El intervalo de confianza $1-\alpha$ por los dos lados para m_X con σ_X^2 desconocida es, como se sabe,

$$\left[\overline{X} - \frac{S_X}{\sqrt{n-1}}\,t, \quad \overline{X} + \frac{S_X}{\sqrt{n-1}}\,t\right]$$

donde t es tal que $F_T(t) = 1-\alpha/2$ siendo T una variable t de Student con $n-1$ grados de libertad. Incorporando los datos del problema obtenemos

- para $\alpha = 0.1$ el intervalo $[8.134, 11.094]$
- para $\alpha = 0.05$ el intervalo $[7.809, 11.419]$
- para $\alpha = 0.01$ el intervalo $[7.097, 12.131]$

b) Si, como es habitual, sólo interesa una cota superior para la varianza, el intervalo de confianza $1-\alpha$ será de la forma $[0, nS_X^2/a]$ donde a es tal que $F_Z(a) = \alpha$, siendo Z una variable χ^2_{n-1}. Incorporando los datos del problema obtenemos

- para $\alpha = 0.1$ el intervalo $[0, 18.044]$
- para $\alpha = 0.05$ el intervalo $[0, 21.565]$
- para $\alpha = 0.01$ el intervalo $[0, 30.937]$

2.– Se quiere estimar la profundidad a que se encuentra roca bajo una capa de arcilla utilizando cierto instrumento sónico. Cada lectura del mismo es una variable aleatoria normal, cuya media teórica μ es la profundidad real (desconocida) y cuyo coeficiente de variación $V_x = 0.2$ se supone constante. Se toman n lecturas del instrumento y se calcula la media muestral $\overline{x}$. Calcular un intervalo de confianza al 90 % de la profundidad a que se encuentra la roca si $n = 16$ y $\overline{x} = 31$ m.

SOLUCIÓN

Sabemos que el intervalo de confianza del 90 % sobre la media de la población se calcula, en la situación que presenta el enunciado, de la siguiente manera

$$0.9 = P\Big[-a < \frac{\overline{x}-\mu}{\sigma/\sqrt{n}} \leq a\Big]$$

donde la variable

$$\frac{\overline{x}-\mu}{\sigma/\sqrt{n}} \equiv N(0,1)$$

Por lo tanto, $a = 1.645$. Operando en la expresión anterior

$$0.9 = P\Big[-a\frac{\sigma}{\sqrt{n}} < \overline{x}-\mu \leq a\frac{\sigma}{\sqrt{n}}\Big]$$

Pero sabemos que el coeficiente de variación, $V_x = \frac{\sigma}{\mu}$, es conocido. Por tanto $\sigma = V_x\mu$. Sustituyendo en la expresión anterior

$$0.9 = P\Big[-a\frac{V_x\mu}{\sqrt{n}} < \overline{x}-\mu \leq a\frac{V_x\mu}{\sqrt{n}}\Big] = P\Big[-\frac{aV_x}{\sqrt{n}} < \frac{\overline{x}}{\mu} - 1 \leq \frac{aV_x}{\sqrt{n}}\Big] = P\Big[1-\frac{aV_x}{\sqrt{n}} < \frac{\overline{x}}{\mu} \leq 1+\frac{aV_x}{\sqrt{n}}\Big] =$$

$$= P\Big[\frac{1}{1+\frac{aV_x}{\sqrt{n}}} < \frac{\mu}{\overline{x}} \leq \frac{1}{1-\frac{aV_x}{\sqrt{n}}}\Big]$$

Luego el intervalo de confianza del 90 % sobre la media poblacional es

$$IC_{90\,\%} = \left[\frac{\overline{x}}{1+\frac{aV_x}{\sqrt{n}}}, \frac{\overline{x}}{1-\frac{aV_x}{\sqrt{n}}}\right]$$

Como $a = 1.645$, $V_x = 0.2$, $n = 16$ y $\overline{x} = 31$ obtenemos operando

$$\boxed{IC_{90\,\%} = [28.644, 33.778]}$$

3.– Para medir la vibración de una sonda de perforación de roca se utilizan dos medidores Doppler, de diferentes fabricantes, que utilizan distintos generadores de rayo laser. Sobre una sonda concreta se realizan 200 ensayos con uno de los medidores, obteniéndose una media de frecuencia $\bar{x}_1 = 3.5 \times 10^3\ Hz$ y una desviación típica de $S_1 = 0.2 \times 10^3\ Hz$. El resultado de 150 ensayos con el segundo medidor da como resultados $\bar{x}_2 = 3.4 \times 10^3\ Hz$ y $S_2 = 0.15 \times 10^3\ Hz$. Para comparar dichos medidores de acuerdo con la normativa se debe calcular un intervalo de confianza $(1-\alpha = 0.99)$ sobre la diferencias de las frecuencias medias poblacionales medidas por ambos aparatos (los aparatos no miden directamente la frecuencia de vibración de la sonda, sino la detección de un cambio Doppler en la frecuencia de la luz coherente dispersada por un objetivo en movimiento, del cual se obtiene una medición resuelta en el tiempo de la velocidad del objetivo). Calcular dicho intervalo silas desviaciones típicas son las verdaderas $(S_1 = \sigma_1,\quad S_2 = \sigma_2)$.

A la vista de los resultados y sin realizar más cálculos, ¿pueden considerarse iguales las medias proporcionadas por ambos aparatos?

SOLUCIÓN

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

$$\boxed{IC_{99\,\%} = [0.0518 \times 10^3, 0.1482 \times 10^3]}$$

Dado que el valor $m_1 - m_2 = 0$ no está incluido en los intervalos de confianza calculados, y dado el valor bastante alto de los tamaños de muestra, podemos concluir que ambos aparatos dan resultados distintos para el mismo fenómeno físico.

4.– La distribución

$$f_X(x) = \alpha\beta\Big[1 + \beta x\Big]^{-(\alpha+1)}, \quad x \geq 0, \quad \alpha > 2, \quad \beta > 0$$

se denomina de Lomax o de Pareto Tipo II. Se utiliza fundamentalmente en problemas de economía, finanzas y similares. En esta distribución se verifica

$$F_X(x) = 1 - [1 + \beta x]^{-\alpha}, \quad E[X] = \frac{1}{\beta(\alpha - 1)}, \quad Var[X] = \frac{\alpha}{\beta^2(\alpha-1)^2(\alpha-2)}$$

Consideremos en particular el consumo bruto de energía eléctrica anual per cápita en núcleos de población medido en kWh. Según diferentes estudios dicha variable aleatoria puede considerarse distribuida de acuerdo con la distribución anterior.

Para calcular el consumo medio en España, X, se han elegido adecuadamente 50 municipios de los 8124 existentes y se ha calculado el consumo anual per cápita en cada uno de ellos. Esta muestra ha proporcionado, entre otros, los siguientes datos:

$$\max\{x_i\} = 33.1621, \quad \min\{x_i\} = 0.00671, \quad \sum_{i=1}^{50} x_i = 112.0198$$

$$\sum_{i=1}^{50} x_i^2 = 1195.6575, \quad \sum_{i=1}^{50} \ln(1 + x_i) = 38.4564, \quad \sum_{i=1}^{50} \ln^2(1 + x_i) = 57.8090$$

Suponiendo que el parámetro $\beta = 1$, calcular los estimadores de α mediante el método de los momentos y mediante el de máxima verosimilitud.

SOLUCIÓN

a) Utilicemos primero el método de los momentos:

$$\hat{m} = \bar{x} = E[x] = \frac{1}{(\hat{\alpha}-1)} \quad \Rightarrow \quad \hat{\alpha} = \frac{1}{\bar{x}} + 1$$

Pero

$$\bar{x} = \frac{\sum_{i=1}^{50} x_i}{50} = 2.240396 \quad \Rightarrow \quad \boxed{\hat{\alpha} = 1.446350}$$

b) La función de verosimilitud es

$$L = \prod_{i=1}^{50} f_X(x_i) = \prod_{i=1}^{50} \alpha \Big[1 + x_i\Big]^{-(\alpha+1)} = \alpha^{50} \prod_{i=1}^{50} \Big[1 + x_i\Big]^{-(\alpha+1)}$$

Tomando logaritmos

$$\ln L = 50 \ln \alpha + \sum_{i=1}^{50} (-(\alpha+1)\ln(1+x_i) = 50 \ln \alpha - (\alpha+1) \sum_{i=1}^{50} \ln(1+x_i)$$

Derivando con respecto a α,

$$\frac{\partial \ln L}{\partial \alpha} = \frac{50}{\alpha} - \sum_{i=1}^{50} \ln(1+x_i)$$

Igualando a cero para hallar el máximo

$$\frac{50}{\hat{\alpha}} = \sum_{i=1}^{50} \ln(1+x_i) \quad \Rightarrow \quad \hat{\alpha} = \frac{50}{\sum_{i=1}^{50} \ln(1+x_i)}$$

y sustituyendo por los valores adecuados

$$\boxed{\hat{\alpha} = \frac{50}{38.4564} = 1.300174}$$

valor obviamente distinto del obtenido por el método de los momentos y que es el que se debería usar.

5.– Un temporal del NW que bate sobre un dique de abrigo puede causar daños (o no) a dicho dique. La probabilidad de que no cause perjuicios económicos (daños de 0 €) es b. Si causa daños, el volumen de los daños D (en €) puede considerarse una variable aleatoria exponencialmente distribuida con parámetro λ. Se desea modelar el coste de los daños sufridos en el dique durante un temporal de NW, a partir de una muestra de los daños experimentados en n temporales del NW $\{d_1, d_2, ..., d_n\}$

a) Calcular, mediante el método de máxima verosimilitud, estimadores puntuales de b y λ.

b) ¿Son insesgados los estimadores calculados en el apartado anterior?

——————SOLUCIÓN——————

a) La variable D (daños producidos en el dique) es obviamente una variable mixta con la siguiente distribución:

$$\begin{cases} P_D(0) = b; & d = 0 \\ \\ f_D(d) = (1-b)\lambda e^{-\lambda d}; & d > 0 \end{cases}$$

Obsérvese que la función de distribución exponencial ha de multiplicarse por $(1-b)$ (que es la probabilidad de que el temporal cause daños) para que las expresiones anteriores representen una distribución adecuada, con suma de probabilidad uno.

Supongamos que tenemos una muestra $d_1, d_2, ..., d_n$ de tamaño n. En esa muestra habrá n_1 valores nulos y n_2 valores no nulos (obviamente $n_1 + n_2 = n$). Vamos a resolver el problema mediante el método de máxima verosimilitud. La función de verosimilitud de la muestra descrita es

$$L = \prod_{i=1}^{n_1} P_D(0) \prod_{i=1}^{n_2} (1-b)\lambda e^{-\lambda d_i} = b^{n_1}(1-b)^{n_2}\lambda^{n_2} e^{\left(-\lambda \sum_{i=1}^{n_2} d_i\right)}$$

Derivando con respecto a b

$$\frac{\partial L}{\partial b} = n_1 b^{n_1-1}(1-b)^{n_2}\lambda^{n_2} e^{\left(-\lambda \sum_{i=1}^{n_2} d_i\right)} - b^{n_1} n_2 (1-b)^{n_2-1}\lambda^{n_2} e^{\left(-\lambda \sum_{i=1}^{n_2} d_i\right)}$$

Igualando a cero y simplificando (nótese que $\lambda > 0$)

$$n_1 \hat{b}^{n_1-1}(1-\hat{b})^{n_2} = \hat{b}^{n_1} n_2 (1-\hat{b})^{n_2-1}$$

y operando resulta

$$\hat{b} = \frac{n_1}{n}$$

lo cual era fácilmente predecible.

Derivando ahora con respecto a λ obtenemos

$$\frac{\partial L}{\partial \lambda} = b^{n_1}(1-b)^{n_2} n_2 \lambda^{n_2-1} e^{\left(-\lambda \sum_{i=1}^{n_2} d_i\right)} - b^{n_1}(1-b)^{n_2}\lambda^{n_2} \sum_{i=1}^{n_2} d_i \;\; e^{\left(-\lambda \sum_{i=1}^{n_2} t_i\right)}$$

Igualando a cero y simplificando, y teniendo en cuenta que $b > 0, (1-b) > 0, \lambda > 0$, resulta

$n_2 = \hat{\lambda} \sum_{i=1}^{n_2} d_i$, por lo que

$$\hat{\lambda} = \frac{n_2}{\sum_{i=1}^{n_2} d_i}$$

es decir, la inversa de la media aritmética de los valores no nulos de la muestra

b) Es evidente que el estimador $\hat{b}$ no es sesgado ya que es el estimador de máxima verosimilitud de la probabilidad de éxito de una variable de Bernouilli, que sabemos que es insesgado.

En el segundo parámetro, $\hat{\lambda}$, obsérvese que lo obtenido es lo mismo que habríamos calculado si hubiéramos eliminado todos los resultados nulos, así como la parte discreta de la distribución. En este caso, el estimador que hemos obtenido es el clásico insesgado de la distribución exponencial habitual, que es insesgado (como sabemos).

———