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 con $\alpha = 0.1, 0.05, 0.01$, así como una cota superior 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.– 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 $kW \cdot h$. 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

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

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.

3.– En el diseño de una obra marítima es fundamental la definición de la altura de ola significante y el periodo de retorno asociado. Para el diseño de un cierto dique se ha estimado que la llegada de temporales sigue una distribución de Poisson con media anual $\nu = 2$ temporales y que la altura máxima de ola en cada temporal, Y, sigue una distribución exponencial con media 2 metros. Se pide calcular la altura de ola significante h_T para que el periodo de retorno asociado sea de 100 años.

Nota: El periodo de retorno T expresado en años de una ola de altura h_T, expresada en metros, es el valor que verifica $P[h \geq h_T] = 1/T$.

Nota 2: Se recuerda que $\displaystyle\sum_{k=0}^{\infty} \frac{z^k}{k!} = e^z$

——————————SOLUCIÓN——————————

Sean N el número de temporales que ocurren en un año y H la altura de ola máxima alcanzada en ese año. Obviamente, $N \equiv P(\nu)$. Por el Teorema de la Probabilidad Total, para todo $h \geq 0$

$$P[H \leq h] = \sum_{n=0}^{\infty} P[H \leq h | N = n]\ P[N = n]$$

Calculemos la distribución de la variable condicionada $H|N = n$. Supuesto que ha sucedido n veces el fenómeno, el hecho de que el máximo nivel alcanzado sea menor o igual que h es la intersección de los sucesos consistentes en que cada una de las n veces el nivel haya sido menor o igual que h; asumiendo que los niveles alcanzados en distintas ocurrencias del fenómeno son independientes, y teniendo en cuenta que estos niveles siguen la misma distribución (exponencial de parámetro $\lambda = 0.5$), se deduce $P[H \leq h|N = n] = P[Y \leq h]^n$ siendo Y una exponencial de parámetro λ, es decir,

$$P[H \leq h|N = n] = (1 - e^{-\lambda h})^n$$

y por lo tanto

$$\begin{aligned} P[H \leq h] &= \sum_{n=0}^{\infty} (1 - e^{-\lambda h})^n e^{-\nu} \frac{(\nu)^n}{n!} = e^{-\nu} \sum_{n=0}^{\infty} \frac{(\nu(1 - e^{-\lambda h}))^n}{n!} \\ &= e^{-\nu} e^{\nu(1-e^{-\lambda h})} = e^{-\nu e^{-\lambda h}}, \quad h \geq 0 \end{aligned}$$

La determinación completa de la distribución de Y es

$$F_H(h) = P[H \leq h] = \begin{cases} e^{-\nu e^{-\lambda h}} & \text{si } h \geq 0 \\ 0 & \text{si } h < 0 \end{cases}$$

Una vez calculada la distribución de Y podemos emplear la definición de periodo de retorno para calcular la altura de ola h_T

$$P[H \geq h_T] = 1/T \Rightarrow P[H \leq h_T] = F_H(h_T) = e^{-\nu e^{-\lambda h_T}} = 1 - 1/T$$

de donde despejando, obtenemos el valor de h_T:

$$h_T = -\frac{1}{\lambda} \ln \left[-\frac{1}{\nu} \ln \left(1 - \frac{1}{T} \right) \right]$$

Sustituyendo los datos del enunciado se obtiene $h_{100} = 10.59$ metros.