CPE (SEGUNDO CURSO)

PRÁCTICA 16 **SOLUCIONES** **(Curso 2024–2025)**

1.– Una empresa de fabricación de cables decide realizar un control de calidad de las bobinas realizando $n = 4$ ensayos destructivos. De esta manera se obtienen 4 medidas experimentales del límite elástico del cable. Se supone que las medidas son independientes y responden a una distribución $N(m, \sigma^2)$, donde σ^2 es desconocida. El control se efectúa mediante un contraste de hipótesis con significación $\alpha = 0.1$ sobre

$$H_o: \; m \geq \; 400, \quad H_1 : < \; 400$$

En una prueba se obtuvo una media muestral de 375 y una desviación típica $S^* = 25$. Se pide:

a) Determinar el estadístico adecuado para el contraste y la correspondiente región de aceptación.

b) Determinar la probabilidad del error Tipo II si el verdadero límite elástico medio fuera de 380.

c) Determinar la probabilidad de rechazar la hipótesis primaria si el verdadero límite elástico medio fuera 405.

d) Supóngase que con la misma significación y el mismo número de datos, pero con otros valores muestrales, se ha obtenido que el error Tipo II para $m = 380$ es 0.08484 y la probabilidad de rechazar la hipótesis primaria para $m = 405$ es 0.29530. Determínese cual fue en este caso la región de aceptación y cuál es el valor de la desviación típica muestral obtenida

Nota: todos los valores de los límites elásticos se indican en MPa

——————SOLUCIÓN——————

a) El contraste de hipótesis que debemos realizar es

$H_o : m \geq 400$

$H_1 : m < 400$

que como sabemos es equivalente, desde un punto de vista operativo a

$H_o : m = 400$

$H_1 : m < 400$

Como la varianza es desconocida utilizaremos el estadístico $t = \frac{\bar{x}-m}{S*/\sqrt{n}} \equiv t_{n-1}$. En nuestro caso $n = 4$. Dado el tipo de contraste, la región crítica estará en la cola inferior de la distribución del estadístico t. Si suponemos que la desviación típica muestral que hemos obtenido es la insesgada, es decir $S^* = 25$, en el supuesto de la hipótesis primaria $t = \frac{\bar{x}-400}{25/\sqrt{4}} \equiv t_3$. Por tanto nuestra regla de decisión, para $\alpha = 0.10$, será rechazar la hipótesis primaria si $t < c$, donde $F_t(c) = 0.10$. De las tablas $c = -1.6378$. De nuestra muestra el estadístico t vale

$t = \frac{375-400}{25/\sqrt{4}} = -2$ y por lo tanto debemos rechazar la hipótesis primaria a este nivel de significación.

b) La probabilidad de error Tipo II es

$$\beta = P[t \geq c|H_1] = P\Big[\frac{\bar{x}-400}{25/\sqrt{4}} \geq c|H_1\Big] = P\Big[\bar{x} \geq 400 + c\frac{25}{\sqrt{4}}|H_1\Big] =$$

$$= P\Big[\bar{x} - m_1 \geq 400 - m_1 + c\frac{25}{\sqrt{4}}|H_1\Big] = P\Big[\frac{\bar{x}-m_1}{25/\sqrt{4}} \geq c + \frac{400-m_1}{25/\sqrt{4}}|H_1\Big] = 1 - F_t\Big(c + \frac{400-m_1}{25/\sqrt{4}}\Big)$$

Para $m_1 = 380$,

$$\beta = 1 - F_t\Big(c + \frac{400-380}{25/\sqrt{4}}\Big) = 1 - F_t\Big(-1.6378 + \frac{400-380}{25/\sqrt{4}}\Big) = 1 - F_t(-0.0378) = F_t(0.0378) = 0.51388$$

c) La probabilidad q de rechazar la hipótesis primaria si el verdadero límite elástico medio fuera 405 es

$$q = P[t < c|m_1 = 405] = P\Big[\frac{\bar{x}-400}{25/\sqrt{4}} < c|m_1 = 405\Big] = P\Big[\frac{\bar{x}-m_1}{25/\sqrt{4}} < c + \frac{400-m_1}{25/\sqrt{4}}|m_1 = 405\Big] =$$

$$= F_t\Big(- 1.6378 + \frac{400-405}{25/\sqrt{4}}\Big) = F_t(-2.0378) = 0.06716$$

d) Nos dicen por una parte que

$$\beta = 0.08484 = P[t \geq c|H_1] = 1 - F_t\Big(c + \frac{400-380}{S^*/\sqrt{4}}\Big)$$

y por otra

$$q = 0.29530 = F_t\Big(c + \frac{400-405}{S^*/\sqrt{4}}\Big)$$

Consecuentemente tenemos estas dos ecuaciones con dos incógnitas (c y S^*):

$$0.91516 = F_t\Big(c + \frac{400-380}{S^*/\sqrt{4}}\Big) = F_t(c + \frac{40}{S^*})$$

$$0.29530 = F_t\Big(c + \frac{400-405}{S^*/\sqrt{4}}\Big) = F_t(c - \frac{10}{S^*})$$

Y de las tablas obtenemos

$$c + \frac{40}{S^*} = 1.8$$

$$c - \frac{10}{S^*} = -0.6$$

de donde

$$S^* = 20.833, \qquad c = -0.12$$

por lo que nuestra región de aceptación es, usando el estadístico $t = \frac{\bar{x}-400}{18.048/\sqrt{4}} = \frac{\bar{x}-400}{9.024} \equiv t_3$,

$$(-0.12, \infty)$$

2.– 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

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

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.

3.– El consumo diario de agua, X, en m^3, en una cierta instalación, sigue una distribución uniforme entre 0 y el consumo máximo a (es decir, $X \equiv U(0,a)$). Se quiere contrastar, con un nivel de significación del 5 %, la hipótesis de que el consumo máximo es de 25 m^3 frente a la alternativa de que es menor que esa cantidad. Para ello, se toma una muestra de diez días, en los que el consumo de agua en m^3 resulta ser: 10, 12, 15, 12, 20, 16, 9, 17, 14 y 11.

a) ¿Qué hipótesis debe ser aceptada?

b) Si el verdadero consumo máximo fuese de 20 m^3, ¿qué probabilidad tendríamos de aceptar la hipótesis primaria?

c) Con estos datos, ¿a qué nivel máximo de significación puede aceptarse la hipótesis primaria?

d) Comentar cómo cambiaría el resultado si la hipótesis alternativa hubiera sido *el consumo máximo es distinto que esa cantidad.*

—————————SOLUCIÓN—————————

a) El test más natural para contrastar $H_0 : a = 25$ frente a $H_1 : a < 25$ es el que proporciona la región crítica de la forma

$$R_C : \text{máx}\{X_1, \cdots, X_n\} < c$$

donde c dependerá del nivel de significación escogido. De hecho ésta sería la región crítica que nos obtendríamos con la aplicación del método de cociente de verosimilitudes (el estimador máximo verosímil del extremo superior a del intervalo es el máximo de la muestra).

Veamos la distribución de la variable aleatoria $Y = \text{máx}\{X_1, \cdots, X_n\}$ siendo $\{X_1, \cdots, X_n\}$ una muestra aleatoria de una variable uniformemente distribuida en el intervalo $[0, a]$. Es claro que $R_Y = (0, a)$ y que para todo y en este intervalo

$$F_Y(y) = P[\text{máx}\{X_1, \cdots, X_n\} \leq y] =$$

$$= P[\{X_1 \leq y\} \cap \{X_2 \leq y\} \cap \cdots \cap \{X_n \leq y\}] =$$

$$= P[X_1 \leq y]P[X_2 \leq y] \cdots P[X_n \leq y] = F_X(y)^n = \frac{y^n}{a^n}$$

ya que $F_X(x) = x/a$ para $x \in [0, a]$.

Como el nivel de significación del test ha de ser $\alpha = 0.05$, y el tamaño muestral es de $n = 10$, se ha de cumplir $0.05 = P_{a=25}[\text{máx}\{X_1, \cdots, X_{10}\} < c] = c^{10}/25^{10}$ y se deduce $c = 18.528$. Obtenemos así el test definido por la región crítica $R_C : \text{máx}\{X_1, \cdots, X_{10}\} < 18.528$.

Como el máximo de nuestra muestra es 20, la muestra cae fuera de la región crítica y conduce a aceptar la hipótesis primaria.

Hay que tener en cuenta que si el máximo de los valores muestrales hubiese resultado mayor que 25, esto sería incompatible tanto con la hipótesis primaria como con la alternativa, independientemente del nivel de significación escogido (ningún valor del parámetro a menor o igual que 25 se correspondería con esos datos).

b) Con el test obtenido en el apartado **a)**, que hemos visto tiene como región crítica $R_C : \text{máx}\{X_1, \cdots, X_{10}\} < 18.528$, y supuesto que el verdadero valor de a es 20, la probabilidad de aceptar sin embargo H_0 se calcula como

$$\beta(a = 20) = P_{a=20}[\text{máx}\{X_1, \cdots, X_{10}\} \geq 18.528] = 1 - \frac{18.528^{10}}{20^{10}}$$

ya que la función de distribución del máximo es en este caso $F_Y(y) = y^{10}/20^{10}$. Se obtiene $\beta = 0.5344$.

c) Nos piden el nivel p de la muestra, es decir, el nivel de significación del test con región crítica $R_C : \text{máx}\{X_1, \cdots, X_{10}\} < c$ lo más grande posible, para el que todavía se acepta la hipótesis primaria, y es claro que esto ocurre precisamente para $c = 20$. Así

$$p = P_{a=25}[\text{máx}\{X_1, \cdots, X_{10}\} < 20] = 20^{10}/25^{10} = 0.10737$$

que es razonablemente alto (a los niveles de significación habituales de 0.01, 0.05, 0.1 la muestra conduce a la aceptación de la hipótesis primaria).

d) Para el contraste de $H_0 : a = 25$ frente a $H_1 : a \neq 25$, el resultado cambiaría sólo en el sentido de que la región crítica sería de la forma

$$R_C : [\text{máx}\{X_1, \cdots, X_{10}\} < c \quad \cap \quad \text{máx}\{X_1, \cdots, X_{10}\} > 25] \,.$$

Los valores del máximo mayores que 25 conducen a rechazar con rotundidad la hipótesis primaria, independientemente del nivel de significación escogido: son incompatibles con $a = 25$. Por esa misma razón el valor de c, para cada nivel de significación escogido, es el mismo que el del contraste del apartado **a)**: el subconjunto $[\text{máx}\{X_1, \cdots, X_{10}\} > 25]$ de la región crítica tiene probabilidad 0 bajo la hipótesis $a = 25$ y por lo tanto no influye en el cálculo de c.

4.– Una compañía especializada en transportes pesados realiza la siguiente oferta a una empresa de construcción de álabes de aerogeneradores que necesita enviar su mercancía a un determinado puerto de mar para embarcarlas. El tiempo de viaje hasta el puerto se pacta en una cifra T_0. Si el camión tarda T_0 el precio acordado es fijo. Ese precio cubre los gastos del transporte, por lo que teóricamente, la empresa de transporte no obtendría ningún beneficio. Llamemos $X = T - T_0$, el tiempo de retraso ($X \geq 0$) o de adelanto ($X < 0$) del camión. Si el camión se retrasa, la empresa de transporte pagaría una multa de $40X^2$. Si por el contrario, el camión se adelanta, la empresa obtendría un pago adicional de $-10X^3$. La empresa sabe que este tiempo X de retraso o adelanto es aleatorio con la siguiente distribución:

$$f_X(x) = \begin{cases} \dfrac{-xe^{-(x/t)^2}}{t^2} & x \leq 0 \\ \dfrac{xe^{-(x/t)^2}}{t^2} & x > 0 \end{cases}$$

Por encargos anteriores, la empresa de transporte tiene una muestra de 12 encargos similares, de donde obtuvo la siguiente muestra, considerada aleatoria, de la variable X : $11.3,\ -3.5,\ -4.5,\ -7.2,\ 3.7,\ 3.5,\ -2.3,\ -0.87,\ 2.2,\ 2.5,\ -5.3,\ -4.1$

Se pide:

a) Determinar un estimador del parámetro t mediante el método de máxima verosimilitud.

b) En el caso de que la fábrica aceptase la oferta, ¿es ésta beneficiosa para la empresa de transporte? Justificar la respuesta.

Nota 1: $n = 12, \quad \sum x_i = -4.57, \quad \sum x_i^2 = 300$

Nota 2: $\displaystyle\int_0^\infty x^2 e^{-(x/t)^2} dx = \frac{t^3\sqrt{\pi}}{4}, \quad \int_0^\infty x^3 e^{-(x/t)^2} dx = \frac{t^4}{2}, \quad \int_0^\infty x^4 e^{-(x/t)^2} dx = \frac{3t^5\sqrt{\pi}}{8}$

SOLUCIÓN

Lo primero que habrá que obtener es un estimador del parámetro t de la distribución. No podemos utilizar el método de los momentos porque la esperanza matemática de X es cero (obvio por simetría de la función). Utilizaremos pues el método de máxima verosimilitud. La función de verosimilitud es

$$L = \prod_{i=1}^{n} f_X(x_i) = \prod_{i=1}^{n} \frac{|x_i| e^{-(x_i/t)^2}}{t^2} = t^{-2n} \prod_{i=1}^{n} \left[|x_i| e^{-(x_i/t)^2} \right]$$

Tomando logaritmos

$$\ln L = -2n\ln(t) + \sum_{i=1}^{n} \ln\left[|x_i|e^{-(x_i/t)^2}\right] = -2n\ln(t) + \sum_{i=1}^{n} \ln|x_i| - \sum_{i=1}^{n}\left[-(x_i/t)^2\right]$$

Derivando con respecto a t

$$\frac{d\ln L}{t} = -\frac{2n}{t} + \frac{2\sum_{i=1}^{n} x_i^2}{t^3}$$

Igualando a cero

$$\frac{2n}{\hat{t}} = \frac{2\sum_{i=1}^{n} x_i^2}{\hat{t}^3}$$

Luego

$$\hat{t} = \sqrt{\frac{\sum_{i=1}^{n} x_i^2}{n}}$$

Con los datos que tenemos, $\hat{t} = 5$

Calculemos la esperanza matemática de los beneficios

$$E[B] = E[-10X^3|X \leq 0] = \int_{-\infty}^{0} \frac{10x^4}{5^2}e^{-(\frac{x}{5})^2}dx = \int_{0}^{\infty} \frac{10x^4}{5^2}e^{-(\frac{x}{5})^2}dx = \frac{30\times 5^3\sqrt{\pi}}{8} = 830.84$$

La esperanza matemática de la multa es, por otro lado

$$E[M] = E[40X^2|X > 0] = \int_{0}^{\infty} \frac{40x^3}{5^2}e^{-(\frac{x}{5})^2}dx = \frac{40}{5^2}\frac{5^4}{2} = 500$$

Luego el trato es beneficioso en media para la empresa, ya que $E[B] > E[M]$

5.– Las ventas de un comercio se producen según un proceso de Poisson de parámetro $\nu = 2$ (ventas/hora). Los ingresos por cada venta se suponen independientes y exponencialmente distribuidos con parámetro $\lambda = (1/60)$ €$^{-1}$.

a) Determínese la distribución de los ingresos en una jornada de 8 horas, su media y su varianza.

b) Calcúlese la distribución de la venta de mayor cuantía realizada en 8 horas.

———————SOLUCIÓN———————

Sea V el número de ventas en un periodo de 8 horas. Obviamente, $V \equiv P(16)$. Sea X_i el ingreso producido por la venta número i, y sabemos que $X_i \equiv EX(1/60)$

a) Si llamamos Y el total de ingresos en una jornada de 8 horas, se verificará que

$$Y = \begin{cases} 0, & \text{si } V = 0 \\ \sum_{i=1}^{v} X_i, & \text{si } V > 0 \end{cases}$$

luego Y es una variable mixta.

$$P[Y=0]=P[V=0]=\frac{\nu^0 e^{-\nu}}{0!}=e^{-\nu}=e^{-16}=1.13\times 10^{-7}$$

Para $Y>0$, Y es una suma de exponenciales, independientes y con el mismo parámetro. Por tanto, si fijamos $V=v$, resulta $Y|V\equiv\Gamma(v,1/60)$. Por tanto

$$f_{Y|V}(y,v)=\frac{\lambda(\lambda y)^{v-1}e^{-\lambda y}}{(v-1)!}$$

Pero

$$P_V(v)=\frac{\nu^v e^{-\nu}}{v!}$$

luego

$$f_Y(y)=\sum_{v=1}^{\infty}\frac{\lambda(\lambda y)^{v-1}e^{-\lambda y}}{(v-1)!}\;\frac{\nu^v e^{-\nu}}{v!}=\frac{e^{-\lambda y\nu}}{y}\sum_{v=1}^{\infty}\frac{(\lambda y\nu)^v}{(v-1)!v!},\quad y\geq 0$$

con $\nu=16$ y $\lambda=1/60,\quad y\geq 0$

La esperanza matemática de Y dado V es, obviamente, $E[Y|V]=v/\lambda$, dado que $P[V=0]$ es muy pequeña y que $Y|V\equiv\Gamma(v,\lambda)$. Por tanto

$$E[Y]=\sum_{v=1}^{\infty}E[Y|V]P_V(v)=\frac{1}{\lambda}\sum_{v=1}^{\infty}vP_V(v)=\frac{1}{\lambda}E[V]=\frac{\nu}{\lambda}=60\times 16=960\text{ €}$$

El cálculo de la varianza es algo más complicado. Como $Var[Y]=E[Y^2]-E^2[Y]$, tenemos que calcular $E[Y^2]$. Partamos de los momento condicionales.

$$E[Y^2|V=v]=Var[Y|V=v]+E^2[Y|V=v]$$

Pero $Y|V=v\sum_{i=1}^{v}X_i$. Luego $E[Y|V=v]=vm_x$ y $Var[Y|V=v]=v\sigma_x^2$. Luego $E[Y^2|V=v]=v\sigma_x^2+v^2m_x^2$. Y por tanto

$$E[Y^2]=E[E[Y^2|V=v]]=E[v\sigma_x^2+v^2m_x^2]=\sigma_x^2 m_v+m_x^2E[V^2]$$

Pero $E[V^2]=\sigma_v^2+m_v^2$ y por lo tanto

$$E[Y^2]=m_v\sigma_x^2+m_x^2(\sigma_v^2+m_v^2)=m_v\sigma_x^2$$

y la varianza será

$$Var[Y]=E[Y^2]-E^2[Y]=m_v\sigma_x^2+\sigma_v^2m_x^2$$

En esta expresión $m_x=60,\quad\sigma_x^2=360,\quad m_v=16,\quad\sigma_v^2=16$. Luego $Var[Y]=11520$ €^{2} y $\sigma_Y=107.3$ €.

b) En este caso, si Z es la venta de mayor cuantía, $Z=max\{X_i\}$. Si fijamos $V=v$,

$$F_{Z|V}(z,v)=[F_X(z)]^v=\left(1-e^{-\lambda z}\right)^v$$

Por tanto

$$F_Z(z)=\sum_{v=0}^{\infty}\left(1-e^{-\lambda z}\right)^v P_V(v)=\sum_{v=0}^{\infty}\left(1-e^{-\lambda z}\right)^v\;\frac{\nu^v e^{-\nu}}{v!}=e^{-\nu}\sum_{v=0}^{\infty}\frac{\left[\nu\left(1-e^{-\lambda z}\right)\right]^v}{v!}=$$

$$=e^{-\nu}e^{\nu\left(1-e^{-\lambda z}\right)}=e^{-e^{-\lambda z}},\quad z\geq 0$$