ENERO-2018.pdf

AzaharaFS

Inferencia Estadística

3º Grado en Matemáticas

Facultad de Ciencias
Universidad de Granada

ENERO 2018

INFERENCIA ESTADÍSTICA.

1. a) Teoría

b) La distancia en km recorrida por dos tipos de coches A y B ~~son~~ con 50 litros de gasolina, sigue, respectivamente una ley $N(\mu_A, 15)$ y $N(\mu_B, 22)$. Se desea encontrar un intervalo de confianza al nivel 0'99 para $\mu_A - \mu_B$, con amplitud máxima de 4 km observando las distancias recorridas con 50 litros de gasolina por un mismo número de coches de cada tipo, elegidos de forma independiente. ¿Cuál es el mínimo número de coches de cada tipo que deben realizar la prueba para obtener dicho intervalo?

X: Distancia en km del coche A con 50 litros de gasolina $\rightarrow N(\mu_A, 15)$

Y: Distancia en km del coche B con 50 litros de gasolina $\rightarrow N(\mu_B, 21)$

$(X_1, \ldots, X_{n_1})$ $(Y_1, \ldots, Y_{n_2})$ m.a.s. de X e Y, $n_1 = n_2 = n$.

Intervalo de confianza para la diferencia de medias de dos poblaciones normales con varianza conocida:

$$\left(\bar{X} - \bar{Y} - z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}},\ \bar{X} - \bar{Y} + z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}\right)$$

Sea $\alpha = 0{,}01 \Rightarrow \alpha/2 = 0{,}005$ $\quad z_{\alpha/2} = z_{0{,}005} \simeq 1'645$

$\sigma_1^2 + \sigma_2^2 = 36$ $\quad$ luego el intervalo nos queda

$$\left(\bar{X} - \bar{Y} - 1'645\frac{\sqrt{36}}{\sqrt{n}},\ \bar{X} - \bar{Y} + 1'645\frac{\sqrt{36}}{\sqrt{n}}\right)$$

Como queremos una amplitud máxima de 4

$$\left(\cancel{\bar{X}} - \cancel{\bar{Y}} + 1'645\frac{\sqrt{36}}{\sqrt{n}}\right) - \left(\cancel{\bar{X}} - \cancel{\bar{Y}} - 1'645\frac{\sqrt{36}}{\sqrt{n}}\right) \leq 4 \Rightarrow 1'645 \cdot 6\left(\frac{1}{\sqrt{n}} + \frac{1}{\sqrt{n}}\right) \leq 4$$

$$\Rightarrow \frac{2}{\sqrt{n}} \leq 0'405268 \Rightarrow \sqrt{n} \geq \frac{2}{0'405268} = 4'935 \Rightarrow n \geq 24'35$$

Como n es natural, necesitamos mínimo 25 coches de cada tipo para realizar la prueba.

2. Teoría

3. Sea $(X_1,\dots,X_n)$ una m.a.s. de una variable aleatoria X con función de densidad $f_\theta(x) = \frac{\sqrt{\theta}}{2} x^{-3/2} \quad x>\theta$

Determinar cuales son los valores de n para los que existe el UMVUE para θ y encontrarlo en tales casos. Calcular el EMV de θ.

En primer lugar buscamos un estadístico suficiente y completo para esta distribución.

$$\chi = \bigcup_\theta \chi_\theta = \bigcup_\theta (\theta,+\infty) = \mathbb{R}^+$$

$$\Theta = \mathbb{R}^+$$

$$f_\theta^n(x_1,\dots,x_n) = \frac{(\sqrt{\theta})^n}{2^n}\left(\prod_{i=1}^n x_i\right)^{-3/2} \cdot I_{[\theta,+\infty)}(X_{(1)})$$

Por el tma de factorización, $X_{(1)}$ es un estadístico suficiente, basta tomar $h(x_1,\dots,x_n) = \left(\prod_{i=1}^n x_i\right)^{-3/2}$ $g_\theta(T(x_1,\dots,x_n)) = \frac{(\sqrt{\theta})^n}{2^n} \cdot I_{[\theta,+\infty)}(X_{(1)})$

Veamos si es completo.

Calculamos la distribución de $X_{(1)}$:

$$F(x) = \int_\theta^x \frac{\sqrt{\theta}}{2} t^{-3/2}\, dt = 1 - \frac{\sqrt{\theta}}{\sqrt{x}} \qquad F_{X_{(1)}} = 1-(1-F(x))^n = 1-\left(\frac{\sqrt{\theta}}{\sqrt{x}}\right)^n$$

$$\Rightarrow f_{X_{(1)}}(x) = \frac{n}{2x}\left(\frac{\sqrt{\theta}}{\sqrt{x}}\right)^n = \frac{n}{2x}\left(\frac{\theta}{x}\right)^{n/2} \qquad x>\theta$$

Para que el estadístico sea completo necesito que si tomo cualquier función medible $g(T)$ del estadístico suficiente con $E_\theta(g(T)=0)$ $\forall \theta \in \Theta$, se cumple que $P_\theta(g(T)=0)=1$.

$$E_\theta(g(T)) = \int_\theta^{+\infty} g(t)\frac{n}{2t}\left(\frac{\theta}{t}\right)^{n/2} dt = \frac{n\cdot\theta^{n/2}}{2}\int_\theta^{+\infty} g(t)\frac{1}{t^{1+n/2}}\,dt = 0 \iff$$

$$\iff \int_\theta^{+\infty} g(t)/t^{1+n/2}\, dt = 0$$ Usando el TFC, sabemos que existe una primitiva, G, de $g(t)/t^{1+n/2}$ que cumple que $G(+\infty)-G(\theta)=0$ y derivando respecto a θ

WELL DONE NATURE
¡BUEN TRABAJO!
TE MERECES UNA PAUSA.
RE-CHARGE
masque nuit relaxant
relaxing sleep mask
my CLARINS
RE-BOOST
masque défatigant
éclair
instant
reviving mask
my CLARINS
¿QUÉ TAL UNA MASCARILLA PARA CUIDAR TU PIEL MIENTRAS DESCANSAS?
my CLARINS
VEGAN FRIENDLY
TUS TRATAMIENTOS CON EXTRACTOS DE FRUTAS Y PLANTAS PARA UNA PIEL SANA Y BONITA
¡REGÁLATELO O PIDE QUE TE LO REGALEN ESTAS NAVIDADES!
Descúbrelo ahora en CLARINS.COM con un 30%* de descuento. Código: WUOLAH1
*Descuento aplicable sobre la gama My Clarins hasta el 28 de febrero de 2022. No acumulable con otras promociones de descuento y precio fidelidad.

Tenemos que $\frac{g(\theta)}{\theta^{1+n/2}} = 0 \Leftrightarrow g(\theta) = 0 \quad \forall \theta \in \Theta$

$\Rightarrow \{t / g(t) = 0\} \supseteq \{t > 0\}$ y tomando probabilidades

$1 \geq P(g(T) = 0) \geq P(T > 0) = 1$. Y por tanto hemos probado que $T = X_{(1)}$ es estadístico suficiente y completo, luego el UMVUE será una función suya insesgada que sea estimador y que tenga momento de segundo orden finito.

Para ver si es UMVUE imponemos insesgadez:

$$E_\theta(h(T)) = \int_\theta^{+\infty} h(t) \frac{n}{2t}\left(\frac{\theta}{t}\right)^{n/2} dt = \theta \Leftrightarrow \frac{n\theta^{n/2}}{2} \int_\theta^{+\infty} h(t)/t^{1+n/2}\, dt = \theta \Leftrightarrow$$

$$\Leftrightarrow \frac{2\,\theta^{1-n/2}}{n} = \int_\theta^{+\infty} h(t)/t^{1+n/2}$$ De nuevo por el TFC tenemos que

$$\frac{h(\theta)}{\theta^{1+n/2}} = \frac{(n-2)\theta^{-n/2}}{n} \Rightarrow h(\theta) = \frac{(n-2)\theta}{n} \Rightarrow h(T) = \frac{(n-2)T}{n}$$ es candidato a UMVUE.

- Claramente $h(T) = \frac{(n-2)T}{n}$ para $n > 2$, ya que como $T \in \mathbb{R}^+$, $h(T) \in \mathbb{R}^+$ cuando $n > 2$
- Veamos que tiene momento de segundo orden finito

$$E_\theta(h(T)^2) = \int_\theta^{+\infty} \left(\frac{n-2}{n}\right)^2 t^2 \frac{n}{2t}\left(\frac{\theta}{t}\right)^{n/2} dt = \frac{(n-2)^2\theta^{n/2}}{2n} \int_\theta^{+\infty} t^{1-n/2}\, dt =$$

$$= \frac{(n-2)^2\theta^{n/2}}{2n}\left[\frac{t^{2-n/2}}{2-n/2}\right]_\theta^{+\infty} < +\infty \Leftrightarrow 2 - n/2 \leq 0 \Leftrightarrow \boxed{n \geq 4}$$

Por tanto existe UMVUE para $n \geq 4$ y es $h(T) = \frac{(n-2)T}{n}$.

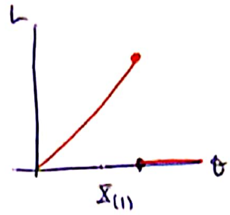
Escaneado con CamScanner

Para calcular el EMV, tenemos que la función de verosimilitud es

$$L_{x_1,\dots,x_n}(\theta) = f_\theta^n(x_1,\dots,x_n) = \frac{(\sqrt{\theta})^n}{2^n}\left(\prod_{i=1}^n x_i\right)^{-3/2} \cdot I_{(0,+\infty)}(X_{(1)})$$

Como no es derivable, ya que ~~...~~ el intervalo de definición depende del parámetro, estudiamos la monotonía de la función (donde no se anula.) que definiremos como

$$g_{x_1,\dots,x_n}(\theta) = \frac{(\sqrt{\theta})^n}{2^n}\left(\prod_{i=1}^n x_i\right)^{-3/2} = \underbrace{(\sqrt{\theta})^n}_{\text{Creciente en } \theta} \underbrace{\frac{\left(\prod_{i=1}^n x_i\right)^{-3/2}}{2^n}}_{\text{No depende de } \theta}$$



$\Rightarrow L_{x_1,\dots,x_n}(\theta)$ es creciente cuando $0<\theta<X_{(1)}$ y se anula en $\theta \geq$, por tanto alcanza su máximo en $X_{(1)}$, así el EMV de θ será $\hat{\theta} = X_{(1)}$.

4. Obtener el test de razón de verosimilitudes de tamaño α para contrastar $H_0: \theta \leq \theta_0$ frente a $H_1: \theta > \theta_0$ basado en una observación de una variable aleatoria con función de distribución exponencial

$$f_\theta(x) = \theta e^{-\theta x} \quad x>0$$

¿Qué tamaño alcanza dicho test?

Sabemos que el TRV es de la forma

$$\varphi(X_1,\dots,X_n) = \begin{cases} 1 & \text{si } \lambda(X_1,\dots,X_n) < c \\ 0 & \text{si } \lambda(X_1,\dots,X_n) \geq c \end{cases} \qquad c \in (0,1]$$

donde $$\lambda(X_1,\dots,X_n) = \frac{\sup_{\theta\in\Theta_0} L_{x_1,\dots,x_n}(\theta)}{\sup_{\theta\in\Theta} L_{x_1,\dots,x_n}(\theta)} \qquad \forall (x_1,\dots,x_n) \in \mathcal{X}^n$$

En nuestro caso: $\left.\begin{array}{l}\Theta_0 = (-\infty, \theta_0] \\ \Theta_1 = (\theta_0, +\infty)\end{array}\right\} \Theta = \Theta_0 \cup \Theta_1 = \mathbb{R}$ $\qquad \mathcal{X} = \mathbb{R}^+$, $\mathcal{X}^n = (\mathbb{R}^+)^n$

Función de verosimilitud asociada a $x\in\mathbb{R}^+$: $L_x(\theta)=\theta e^{-\theta x}$ $\quad \theta\in\mathbb{R}$

$\ln L_x(\theta)=\ln(\theta)-\theta x \longrightarrow \dfrac{\partial \ln L_x(\theta)}{\partial\theta}=\dfrac{1}{\theta}-x=0 \Rightarrow \hat\theta=\dfrac{1}{x}$

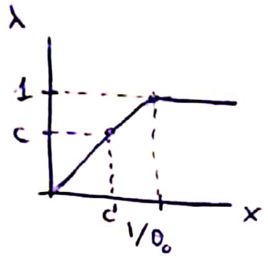
(como tiene un único extremo, es máximo)

$$\lambda(x)=\frac{\sup_{\theta\le\theta_0} L_x(\theta)}{\sup_{\theta\in\mathbb{R}} L_x(\theta)}=\begin{cases}\dfrac{L_x(1/x)}{L_x(1/x)}=1 & \text{si } 0<\frac{1}{x}\le\theta_0\\[2mm] \dfrac{L_x(\theta_0)}{L_x(1/x)}=x\theta_0\cdot e^{1-\theta_0 x} & \text{si } 1/x>\theta_0\end{cases}$$

Estudiemos cómo se comporta $\lambda(x)$:

$1/x>\theta_0 \longrightarrow \ln\lambda(x)=\ln(x\theta_0)+1-\theta_0 x \longrightarrow \dfrac{\partial\ln\lambda(x)}{\partial x}=\dfrac{\theta_0}{x\theta_0}-\theta_0=\dfrac{1}{x}-\theta_0>0$

$\Rightarrow \lambda(x)$ crece en $1/x>\theta_0$ y para $\propto$ $1/x\le\theta_0$ permanece constante



$$\lambda(x)=\begin{cases}1 & \text{si } 0<\frac{1}{\theta_0}\le x\\ x\theta_0\cdot e^{1-\theta_0 x} & \text{si } 1/\theta_0>x\end{cases}$$

$$\Rightarrow \varphi(x)=\begin{cases}1 & \text{si } x<c' \quad (0<c'\le 1/\theta_0)\\ 0 & \text{si } x\ge c'\end{cases}$$

Para calcular c' usamos el tamaño.

- Tamaño: $\sup_{\theta\le\theta_0} E_\theta(\varphi(X))$

$$\alpha=\sup_{\theta\le\theta_0}E_\theta(\varphi(X))=\sup_{\theta\le\theta_0} 1\cdot P_\theta(X<c')=\sup_{\theta\le\theta_0}\int_0^{c'}\theta e^{-\theta x}\,dx=$$

$$=\sup_{\theta\le\theta_0}\left[-e^{-\theta x}\right]_0^{c'}=\sup_{\theta\le\theta_0}\left(-e^{-\theta c'}+1\right)=1-e^{-\theta_0 c'}=\alpha\Rightarrow$$

$$\Rightarrow \ln(\alpha-1)=-\theta_0 c'\Rightarrow c'=-\frac{\ln(\alpha-1)}{\theta_0} \quad //$$

pero $c'=\underbrace{\dfrac{-\ln(\alpha-1)}{\theta_0}}_{\ge 0\ (*)}\in(0,1/\theta_0] \Leftrightarrow -\dfrac{\ln(\alpha-1)}{\theta_0}\le 1/\theta_0 \Leftrightarrow -\ln(\alpha-1)\le 1 \Leftrightarrow$

$\Leftrightarrow \quad \underline{\alpha\ge 1+e^{-1}}$

(*) Como $\alpha\in(0,1)\Rightarrow(1-\alpha)\in(0,1)\Rightarrow \ln(\alpha-1)<0$
y $\theta_0>0$

Por tanto para tamaño $\alpha \geq e^{-1}+1$, el TRV será:

$$\varphi(\underline{x}) = \begin{cases} 1 & \text{si } x < -\dfrac{\ln(\alpha-1)}{\theta_0} \\ 0 & \text{si } x \geq -\dfrac{\ln(\alpha-1)}{\theta_0} \end{cases}$$

6.b) A partir de los siguientes datos sobre el número de visitas a un durante 100 días, contrasta si el nº medio de visitas diarias sigue una distribución de Poisson de parámetro 2.

Nº visitas	0	1	2	3	4	5	6	7
Nº días	29	51	60	30	19	9	1	1

Se trata de un problema de bondad de ajuste, en el que la variable aleatoria observada y las hipótesis a contrastar son:

X: Nº de visitas diarias $\qquad n=200$

$$\begin{cases} H_0: X \longrightarrow P(2) \\ H_1: X \not\longrightarrow P(2) \end{cases}$$

Para poder aplicar el test de Pearson tenemos que particionar el conjunto de valores.

N_i: Número de individuos en cada categoría según los datos (en la categoría A_i: $i=1,\dots,k$)

np_i^0: Número de individuos esperados bajo H_0 (en cada categoría A_i)

Particionamos el conjunto de valores de la distribución hipotética ($\mathbb{N} \cup \{0\}$) en $A_1,\dots,A_8$:

A_i	0	1	2	3	4	5	6	7
prob. esperada p_i^0	$P(X=0)$ 0,1353	$P(X=1)$ 0,2707	$P(X=2)$ 0,2707	0,1804	0,0902	0,0361	0,0120	0,0045
Frec. esperada $200p_i^0$	27,0671	54,1341	54,1341	36,0894	18,0447	7,2179	2,406	0,9068

Como para las categorías $A_i \geq 6$, $200p_i < 5$, y no cumple los requisitos para que el test sea aceptable. Solucionamos esto agrupando las tres últimas categorías para poder aplicar el test.

A_i	0	1	2	3	4	≥ 5
Frec. esperada $200p_i$	27,0671	54,1341	54,1341	36,0894	18,0447	10,5306
Frec. obtenida n_i	29	51	58	39	20	8

$$\chi^2_{exp} = \sum_{i=1}^{6} \frac{(n_i - p_i n)^2}{n p_i} = \frac{(29-27{,}0671)^2}{27{,}0671} + \dots + \frac{(8-10{,}5306)^2}{10{,}5306} = 2{,}06947$$

Y com $\chi^2(N_1, \dots, N_6) \underset{H_0}{\simeq} \chi^2(5)$

$$p\text{-nivel} = P_{H_0}(\chi^2(N_1,\dots,N_6) > 2{,}06947) \in (0'75, 0'8)$$

Se concluye que los datos no proporcionan evidencia para rechazar la hipótesis de que el número de accidentes diarios sigue una distribución $P(2)$.