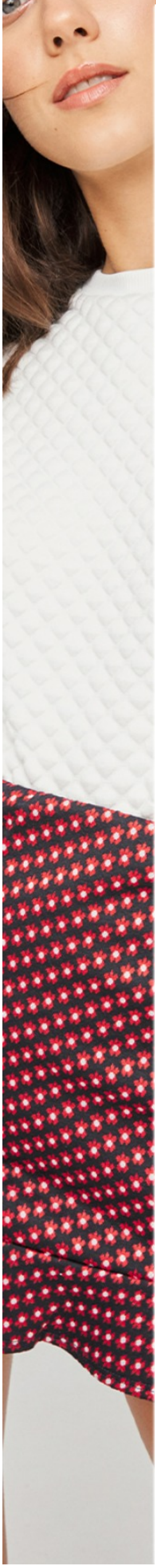
SEPTIEMBRE-2015.pdf

AzaharaFS

Inferencia Estadística

3º Grado en Matemáticas

Facultad de Ciencias
Universidad de Granada



SEPTIEMBRE 2015

INFERENCIA ESTADÍSTICA

1.- Sean $(X_1, \dots, X_n)$, $(Y_1, \dots, Y_m)$ muestras independientes de poblaciones normales con medias μ_1, μ_2 y varianzas σ_1^2, σ_2^2. Sean $\bar{X}, \bar{Y}, S_1^2, S_2^2$ las medias y cuasivarianzas muestrales respectivamente. Partiendo de la independencia de los vectores $(\bar{X}, \bar{Y})$, (S_1^2, S_2^2), deducir la distribución de la siguiente variable, especificando y justificando todos los pasos:

$$\frac{(\bar{X}-\bar{Y})-(\mu_1-\mu_2)}{\sqrt{\frac{\sigma_1^2}{n}+\frac{\sigma_2^2}{m}}\sqrt{\frac{(n-1)S_1^2}{\sigma_1^2}+\frac{(m-1)S_2^2}{\sigma_2^2}}}\sqrt{n+m-2}$$

$X \to N(\mu_1, \sigma_1^2)$

$Y \to N(\mu_2, \sigma_2^2)$

Sabemos que $\bar{X} \to N(\mu_1, \sigma_1^2/n)$ y $\bar{Y} \to N(\mu_2, \sigma_2^2/m)$. Por el lema de Fisher tenemos que los estadísticos $\bar{X}$ e $\bar{Y}$ son independientes, además. Además $\bar{X}-\bar{Y} \to N\left((\mu_1-\mu_2), \frac{\sigma_1^2}{n}+\frac{\sigma_2^2}{m}\right)$

tipificando
$$\frac{(\bar{X}-\bar{Y})-(\mu_1-\mu_2)}{\sqrt{\frac{\sigma_1^2}{n}+\frac{\sigma_2^2}{m}}} \to N(0,1)$$

Por otro lado $\frac{(n-1)S_1^2}{\sigma_1^2} \to \chi^2(n-1)$ y $\frac{(m-1)S_2^2}{\sigma_2^2} \to \chi^2(m-1)$

Como por el lema de Fisher S_1^2 y S_2^2 son independientes, podemos usar la reproductividad de la χ^2, luego

$$\frac{(n-1)S_1^2}{\sigma_1^2}+\frac{(m-1)S_2^2}{\sigma_2^2} \to \chi(m+n-2)$$

Sabemos que si A y B son independientes, $A \to N(0,1)$, $B \to \chi^2(n) \Rightarrow T = \frac{A}{\sqrt{B/n}} \to t(n)$

Tomando $A = \frac{(\bar{X}-\bar{Y})-(\mu_1-\mu_2)}{\sqrt{\frac{\sigma_1^2}{n}+\frac{\sigma_2^2}{m}}}$ y $B = \frac{(n-1)S_1^2}{\sigma_1^2}+\frac{(m-1)S_2^2}{\sigma_2^2}$

nos queda

$$\frac{A}{\sqrt{B/_{n+m-2}}} = \frac{(\bar{X}-\bar{Y})-(\mu_1-\mu_2)}{\sqrt{\frac{\sigma_1^2}{n}+\frac{\sigma_2^2}{m}}\sqrt{\frac{(n-1)S_1^2}{\sigma_1^2}+\frac{(m-1)S_2^2}{\sigma_2^2}}}\sqrt{n+m-2} \longrightarrow t(m+n-2)$$

2.- a) Teoría

b) Teoría

c) Sea $(X_1,\dots,X_n)$ una m.a.s. de una variable X con función de densidad

$$f_\theta(x)=\frac{e^x}{e^\theta-1} \quad 0<x<\theta$$

i) Encontrar un estadístico suficiente y completo, probando que lo es.

ii) Encontrar, si existe, un UMVUE para e^θ.

iii) Encontrar el EMV de e^θ, justificando su obtención ¿Es insesgado este estimador? Justifícalo.

i) Ya que $\mathcal{X}_\theta=(0,\theta)$ depende de θ, la familia no es exponencial uniparamétrica, luego para buscar un estadístico suficiente vamos a utilizar el Teorema de Factorización de Neyman-Fisher.

$$f_\theta^n(x_1,\dots,x_n)=\frac{e^{\sum_{i=1}^n x_i}}{(e^\theta-1)^n} \qquad (x_1,\dots,x_n)\in\mathcal{X}^n=(\mathbb{R}^+)^n$$

Como $\mathcal{X}=\bigcup_{\theta\in\Theta}\mathcal{X}_\theta=\bigcup_{\theta\in\mathbb{R}^+}(0,\theta)=\mathbb{R}^+ \Rightarrow \mathcal{X}^n=(\mathbb{R}^+)^n$, siempre se cumple que $x_i>0$, por tanto $f_\theta^n(x_1,\dots,x_n)$ la podemos expresar como

$$f_\theta^n(x_1,\dots,x_n)=\frac{e^{\sum_{i=1}^n x_i}}{(e^\theta-1)^n}\cdot I_{(-\infty,\theta)}(X_{(n)})$$

y tomando $h(x_1,\dots,x_n)=e^{\sum_{i=1}^n x_i}$ y $g_\theta(T(x_1,\dots,x_n))=\frac{I_{(-\infty,\theta)}(X_{(n)})}{(e^\theta-1)^n}$

tendríamos que $T=X_{(n)}$ es suficiente.

Para ver que es completo veamos que si tengo cualquier función medible g tal que

$E_\theta(g(T))=0 \quad \forall \theta \in \Theta = \mathbb{R}^+ \Rightarrow P_\theta(g(T)=0)=1 \quad \forall \theta \in \Theta$

Así, vamos a calcular primero la función de densidad del estadístico:

$$F_\theta(x) = \int_0^x \frac{e^t}{e^\theta-1}dt = \frac{1}{e^\theta-1}\int_0^x e^t dt = \frac{e^x-1}{e^\theta-1} \quad \text{con } 0<x<\theta$$

$$\Rightarrow F_\theta^T(x) = (F_\theta(x))^n = \left(\frac{e^x-1}{e^\theta-1}\right)^n \rightarrow f_\theta^T(x) = n\left(\frac{e^x-1}{e^\theta-1}\right)^{n-1}\cdot\frac{e^x}{e^\theta-1} = \frac{ne^x(e^x-1)^{n-1}}{(e^\theta-1)^n}$$

$$E_\theta(g(T)) = \int_0^\theta g(t)f_\theta^T(t)dt = \int_0^\theta g(t)\cdot\frac{ne^t(e^t-1)^{n-1}}{(e^\theta-1)^n}dt = \underbrace{\frac{n}{(e^\theta-1)^n}\int_0^\theta g(t)e^t(e^t-1)^{n-1}dt}_{\overset{u}{0}}$$

$\Leftrightarrow \int_0^\theta g(t)e^t(e^t-1)^{n-1}dt = 0$ Por el TFC sabemos que $g(t)e^t(e^t-1)^{n-1}$ tiene primitiva G cumpliendo que $G(\theta)-G(0)=0$. Derivando respecto de θ, tenemos que $g(\theta)\underbrace{e^\theta(e^\theta-1)^{n-1}}_{>0}=0 \Leftrightarrow g(\theta)=0 \quad \forall \theta \in \Theta$

Hemos demostrado que $\{t<\theta\} \subseteq \{t / g(t)=0\}$, y tomando probabilidades tenemos que $1 \geq P(g(T)=0) \geq P(T<\theta) = 1$

$\Rightarrow \underline{P(g(T)=0)=1}$ Por tanto $T=X_{(n)}$ es completo.

ii) UMVUE?

Sabemos que el UMVUE se puede calcular mediante una función de un estadístico suficiente y completo, luego sea $h(T)$ dicha función, imponemos la insesgadez.

$$E_\theta(h(T)) = \int_0^\theta h(t) f_{X_{(n)}}(t)\,dt = \frac{n}{(e^\theta-1)^n}\int_0^\theta h(t)e^t(e^t-1)^{n-1}\,dt = e^\theta \Leftrightarrow$$

$$\Leftrightarrow \int_0^\theta h(t)e^t(e^t-1)^{n-1}\,dt = \frac{e^\theta(e^\theta-1)^n}{n}$$ Aplicando de nuevo el TFC

existe H primitiva de $h(t)e^t(e^t-1)^{n-1}$ tal que $H(\theta)-H(0)=0$,

derivando respecto a θ. $h(\theta)e^\theta(e^\theta-1)^{n-1} = e^{2\theta}(e^\theta-1)^{n-1} + \frac{e^\theta(e^\theta-1)^n}{n}$

$\Rightarrow h(\theta) = e^\theta + \frac{1}{n}$ luego

$h(T) = e^T + \frac{1}{n}$ es candidato a UMVUE

- Veamos que es estimador de e^θ, es decir que toma valores en $\mathbb{R}^+$. Como $T>0 \Rightarrow h(T)>0 \Rightarrow h(T)\in\mathbb{R}^+$, luego sí es estimador
- Veamos por último si tiene momento de segundo orden finito.

$$E_\theta(h(T)^2) = \int_0^\theta h(t)^2 f_{X_{(n)}}(t)\,dt < +\infty$$ ya que tenemos

la integral en un recinto finito ~~de una combinación de exponenciales~~ y $n>0$.

$\Rightarrow h(T) = e^T + \frac{1}{n}$ es UMVUE para e^θ.

iii) Función de verosimilitud:

$$L_{x_1,\dots,x_n}(\theta) \begin{cases} 0 & \text{si } \theta < \max x_i \\ \dfrac{e^{\sum_{i=1}^n x_i}}{(e^\theta-1)^n} & \text{si } \theta > \max x_i \end{cases}$$




Para encontrar el EMV, vamos a buscar el máximo de la función de verosimilitud.

Si $0<\max x_i<\theta$, $L_{x_1,\dots,x_n}(\theta) = \dfrac{e^{\sum_{i=1}^n x_i}}{(e^\theta-1)^n}$

$\ln L_{x_1,\dots,x_n}(\theta) = \sum_{i=1}^n x_i - n\ln(e^\theta-1) \rightarrow \dfrac{\partial \ln L_{x_1,\dots,x_n}(\theta)}{\partial\theta} = -\dfrac{ne^\theta}{e^\theta-1} < 0$ con $0<\max x_i<\theta$

$\Rightarrow$ cuando $0<X_{(n)}<\theta$, $L_{x_1,\dots,x_n}(\theta)$ decrece, por tanto alcanza su máximo en el mínimo valor de θ, es decir ~~$X_{(n)}$~~ $\max x_i$, por tanto $\hat{\theta} = X_{(n)}$. Como nos piden el EMV para e^θ, por el Teorema de invarianza de Zhena tenemos que

$\widehat{e^\theta} = e^{\hat{\theta}} = e^{X_{(n)}}$

¿Es insesgado?

Como ya demostramos que $h(T) = e^T(e^T-1)^{n-2} + \dfrac{(e^T-1)^{n-1}}{n}$ es UMVUE para e^θ, con $T=X_{(n)}$. Como el UMVUE para e^θ es único, entonces se tiene que $\widehat{e^\theta} = e^{X_{(n)}}$ no es UMVUE para e^θ, pues en general $h(T) \neq e^{X_{(n)}}$. Por tanto, tampoco es insesgado, ya que si lo fuera, sería un estimador función del UMVUE para e^θ, contradiciendo así su unicidad.

~~3.- Sea $(X_1,\dots,X_n)$~~ Teoría

~~b) A partir del test anterior~~

Ej de Febrero 2015 (No sé si está bien)

b) El nº de clientes que visitan una oficina sigue una distribución de Poisson con parámetro λ, $P(\lambda)$, y el nº de visitas al día es independiente un día de otro. Para contrastar que el número medio de visitas por día es 0'5 frente a que es 0'6, se hace un estudio durante 10 días y el número total de visitas es 12. Plantear y resolver el contraste de hipótesis adecuada adecuada a la hipótesis con el test más potente de tamaño 0,005 y ver si se acepta o no la hipótesis nula, H_0.

Utilizaremos el test de máxima potencia o test de Neyman-Pearson, con tamaño 0,005.

$X \equiv$ Nº de clientes que visitan una oficina $\quad X \rightarrow P(\lambda)$

$$f_\lambda(x) = \frac{e^{-\lambda}\lambda^x}{x!}$$

$(X_1,\dots,X_{10}) \quad \sum_{i=1}^{10} X_i = 12$

$$\begin{cases} H_0: \lambda = 0'5 \\ H_1: \lambda = 0'6 \end{cases} \qquad \chi_0 = \chi_1 = \mathbb{N}\cup\{0\} \Rightarrow \chi^n = (\mathbb{N}\cup\{0\})^n \qquad \Theta = \Theta_0\cup\Theta_1 = \{0,5;\ 0,6\}$$

Para el test de Neyman-Pearson necesitaremos calcular

$$f_j^n(x_1,\dots,x_n) = \frac{e^{-\lambda_j n}\lambda_j^{\sum_{i=1}^n x_i}}{\prod_{i=1}^n x_i!} \quad (x_1,\dots,x_n)\in\chi^n \quad j=1,0 \quad \text{y } f_0^n(x_1,\dots,x_n)\neq 0 \ \forall (x_1,\dots,x_n)\in\chi^n$$

Por tanto podemos construir el test de Neyman-Pearson de la siguiente forma:

$$\varphi(x_1,\dots,x_n) = \begin{cases} 1 & \text{si } \lambda(x_1,\dots,x_n) > K \\ \gamma & \text{si } \lambda(x_1,\dots,x_n) = K \\ 0 & \text{si } \lambda(x_1,\dots,x_n) < K \end{cases}$$

donde $\lambda(x_1,\dots,x_n) = \dfrac{f_1^n(x_1,\dots,x_n)}{f_0^n(x_1,\dots,x_n)}$

$K \in \mathbb{R}^+\cup\{0\}$

y $\gamma(x_1,\dots,x_n) = \gamma \in [0,1]$

$$\lambda(x_1,\dots,x_n) = \frac{f_1^n(x_1,\dots,x_n)}{f_0^n(x_1,\dots,x_n)} = e^{n(\lambda_0-\lambda_1)} \lambda_1^{\sum_{i=1}^n x_i} \lambda_0^{\sum_{i=1}^n x_i} \quad \text{con } (x_1,\dots,x_n)\in\mathcal{X}^n$$

Como $n=10$ y $\sum_{i=1}^{10} x_i = 12$ nos queda $\lambda(x_1,\dots,x_n) = e^{10(\lambda_0-\lambda_1)}\lambda_1^{12}\lambda_0^{12}$
$= e^{10(\lambda_0-\lambda_1)}(\lambda_1\lambda_0)^{12}$

Le damos a K el valor que toma $\lambda(x_1,\dots,x_n)$.

Como $\lambda(x_1,\dots,x_n) = K = e^{10(\lambda_0-\lambda_1)}(\lambda_1\lambda_0)^{12}$ $\forall (x_1,\dots,x_n)\in\mathcal{X}^n$ el test de Neyman-Pearson nos queda

$\varphi(x_1,\dots,x_n) = \gamma$ Determinamos γ con el tamaño

$\cancel{E_{\theta_0}(\varphi(X_1,\dots,X_n)) = \gamma P_{\theta_0}}$ Como $\forall (x_1,\dots,x_n)\in\mathcal{X}^n$ $\varphi(x_1,\dots,x_n)=\gamma$ el tamaño de nuestro test será γ, es decir $E_{\theta_0}(\varphi(X_1,\dots,X_n)) = 0{,}005 = \gamma$ y por tanto el TMP de tamaño 0,005 será $\varphi(x_1,\dots,x_n) = 0{,}005$ $\forall (x_1,\dots,x_n)\in\mathcal{X}^n$

Es decir, ~~rechazaremos~~ aceptaremos H_0 con probabilidad 0,995 (o ~~aceptamos~~ rechazamos con probabilidad 0,005). Los datos no dan evidencia para aceptar H_0.

4.- Teoría

5.- a) Teoría

b) A partir de una muestra de 200 empleados, se ha construido la siguiente tabla de frecuencias conjunta de calidad de vida (CV) y Rendimiento de Trabajo (RT). Se desea contrastar a nivel 0,05 si los RT son independientes de la CV de los empleados.

CV \ RT	Poca	Media	Mucha
Bajo	62	10	2
Medio	14	38	12
Alto	2	9	51

Se trata de contrastar la independencia de las variables "Calidad de vida" y "Rendimiento de trabajo". Ambas variables categóricas:

$X \equiv$ Calidad de vida

$Y \equiv$ Rendimiento de trabajo

$Y_1 \equiv$ "Bajo" $Y_2 \equiv$ "Medio" $Y_3 \equiv$ "Alto"

$X_1 \equiv$ "Poca" $X_2 \equiv$ "Media" $X_3 \equiv$ "Mucha"

H_0: X, Y independientes o equivalentemente

$$P(X = X_i, Y = Y_j) = P(X = X_i)P(Y = Y_j) \quad i = 1,2,3 \quad j = 1,2,3$$

Debemos estimar las frecuencias esperadas bajo H_0 por máxima verosimilitud $e_{ij} = \frac{n_{i.} n_{.j}}{n}$

RT \ CV	Poca	Media	Mucha	$n_{i.}$
Bajo	62 28,86	10 21,09	2 24,05	74
Medio	14 24,96	38 18,24	12 20,8	64
Alto	2 0,62	9 17,67	51 20,15	62
$n_{.j}$	78	57	65	200 = n

Como todas las frecuencias esperadas son mayores que 2 y que 5, el test χ^2 de independencia es fiable.

$$\chi^2_{exp} = \sum_{i=1}^{3}\sum_{j=1}^{3} \frac{(n_{ij} - e_{ij})}{e_{ij}} = \frac{(62-28,86)^2}{28,86} + \frac{(10-21,09)^2}{21,09} + \dots + \frac{(51-20,15)^2}{20,15} =$$

$$= 38,05473 + 5,83158 + 20,21632 + 4,8126 + 21,40667 + 3,72308 +$$

$$+ 3,07161 + 4,25405 + 47,23189 = 148,60253$$

$$\varphi(N_{ij}) = \begin{cases} 1 & \chi^2_{exp} \geq \chi^2_{4;0,05} \\ 0 & \chi^2_{exp} < \chi^2_{4;0,05} \end{cases}$$

Como $\chi^2_{4;0,05} = 9,4877$

Como $\chi^2_{exp} > \chi^2_{4;0,05}$ No se rechaza la hipótesis H_0 a nivel de significación 0,05. Los datos no dan evidencia de que existe relación entre la calidad de vida y el rendimiento de trabajo.