2.º Teste 25 Jun 2011

Código E

Grupo I

1. $n=10$, $(X_1, X_2, \dots, X_{10})$ aa, $X \sim N(\mu, \sigma=2)$

(a) Função Verosimilhança:

$$\mathcal{L}(\mu \mid x_1 \dots x_n) = f_{X_1 \dots X_n}(x_1 \dots x_n)$$

$$\underset{\text{indep}}{\overset{X_i \text{ são}}{=}} \prod_{i=1}^{n} f_{X_i}(x_i)$$

$$\underset{\text{id}}{\overset{X_i \text{ são}}{=}} \prod_{i=1}^{n} f_X(x_i) = \prod_{i=1}^{n} \left(\frac{1}{8\pi}\right)^{1/2} \exp\left\{-\frac{1}{2}\frac{(x_i-\mu)^2}{4}\right\}$$

$$= \left(\frac{1}{8\pi}\right)^{n/2} \exp\left\{-\frac{1}{2}\sum_{i=1}^{n}\frac{(x_i-\mu)^2}{4}\right\}$$

Função log-Verosimilhança:

$$\ell(\mu \mid x_1 \dots x_n) = \log_e \mathcal{L}(\mu \mid x_1 \dots x_n)$$

$$= \log_e \left\{\left(\frac{1}{8\pi}\right)^{n/2} \exp\left\{-\frac{1}{2}\sum_{i=1}^{n}\frac{(x_i-\mu)^2}{4}\right\}\right\}$$

$$= -\frac{n}{2}\log(8\pi) - \frac{1}{8}\sum_{i=1}^{n}(x_i-\mu)^2$$

Maximização da função log-Verosimilhança:

$$\left.\frac{d}{d\mu}\ell(\mu \mid x_1 \dots x_n)\right|_{\mu=\hat{\mu}} = 0 \;\wedge\; \left.\frac{d^2}{d\mu^2}\ell(\mu \mid x_1 \dots x_n)\right|_{\mu=\hat{\mu}} < 0$$

INSTITUTO SUPERIOR TÉCNICO

(i) $\left.\dfrac{d}{d\mu} \ell(\mu|x_1 \dots x_n)\right|_{\mu=\hat{\mu}} = 0$

$(\Leftrightarrow) \quad \left.\dfrac{1}{4} \sum_{i=1}^{n} (x_i - \mu)\right|_{\mu=\hat{\mu}} = 0 \quad (\Leftrightarrow) \quad n\hat{\mu} = \sum_{i=1}^{n} x_i$

$(\Leftrightarrow) \quad \hat{\mu} = \dfrac{\sum^{n} x_i}{n} \quad (\Leftrightarrow) \quad \hat{\mu} = \bar{x}$

(ii) $\left.\dfrac{d^2}{d\mu^2} \ell(\mu|x_1 \dots x_n)\right|_{\mu=\hat{\mu}} = \left.-\dfrac{1}{4} \sum_{i=1}^{n} 1\right|_{\mu=\hat{\mu}} =$

$-\dfrac{n}{4} < 0$, já que $n=10$

Logo o Estimador de máximo Verosimilhança de μ é $\dfrac{\sum_{i=1}^{10} X_i}{10} = \bar{X}$

(b) Seja $T_1 = \bar{X}$ e $T_2 = \dfrac{X_1 + 3X_{10}}{4}$ estimadores de μ

$e_\mu(T_1|T_2) = \dfrac{EQM(T_2)}{EQM(T_1)}$

onde

$EQM(T_1) = Var(\bar{X}) + (E(\bar{X}) - \mu)^2$

$= \dfrac{var(X)}{10} = \dfrac{4}{10} = 0.4$

INSTITUTO SUPERIOR TÉCNICO

Uma vez que:

$$\text{(i)}\quad E(\bar{X}) = E\left(\frac{\sum X_i}{10}\right) = \sum_{i=1}^{10} \frac{E(X_i)}{10} \overset{X_i \text{ são id}}{=} \frac{10\,E(X)}{10}$$

$$= E(X) = \mu$$ i.e. $\bar{X}$ é estimador centrado de μ.

$$\text{(ii)}\quad Var(\bar{X}) = Var\left(\frac{\sum X_i}{10}\right) \underset{\boxed{X_i \text{ são indep}}}{=} \frac{\sum^{10} Var(X_i)}{10^2} \overset{\boxed{X_i \text{ são id}}}{=} \frac{10 Var(X)}{10^2}$$

$$= \frac{Var(X)}{10} = \frac{4}{10} = 0.4$$

$$EQM(T_2) = Var\left(\frac{X_1+3X_{10}}{4}\right) + \left(E\left(\frac{X_1+3X_{10}}{4}\right) - \mu\right)^2$$

$$\overset{\boxed{X_1 \perp\!\!\!\perp X_{10}}}{=} \frac{var(X_1) + 3^2 Var(X_{10})}{4^2} =$$

$$\overset{\boxed{X_1 \text{ e } X_{10} \text{ são id}}}{=} \frac{4 + 9 \times 4}{16} = \frac{40}{16} = 2.5$$

Uma vez que

$$E\left[\frac{X_1+3X_{10}}{4}\right] = \frac{E(X_1) + 3\,E(X_{10})}{4} \overset{\boxed{X_1 \text{ e } X_{10} \text{ são id}}}{=}$$

$$= \frac{\mu + 3\mu}{4} = \mu$$ i.e. T_2 é estimador centrado de μ

logo $e_{f}(T_1|T_2) = \frac{2.5}{0.4} = 6.25 > 1$

i.e. $EQM(T_2) > EQM(T_1)$ e por isso T_1 é mais eficiente que T_2.

2. Seja $X = \begin{cases} 1, & \text{cliente concorda que o chá cura as dores de cabeça} \\ 0, & \text{caso contrário} \end{cases}$

e $(X_1, X_2, \dots, X_{250})$ é uma amostra aleatória da população X, onde $X \sim Bern(p)$,

$p = P(X=1)$

Sabe-se que $(x_1, x_2, \dots, x_{250})$ é tal que

$$\sum_{i=1}^{250} x_i = 195$$

(a) <u>1. Variável fulcral</u>:

Pelo TLC (aplicável pq $n = 250 \geq 30$,

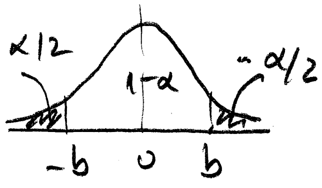
$E(X) = p$, $Var(X) = p(1-p)$ existem)

$$\frac{\bar{X} - p}{\sqrt{\frac{p(1-p)}{n}}} \overset{a}{\sim} N(0,1)$$ e ainda é possível

provar que, usando $\hat{var}(\bar{X}) = \frac{\bar{X}(1-\bar{X})}{n}$ se obtém

$$T = \frac{\bar{X} - p}{\sqrt{\frac{\bar{X}(1-\bar{X})}{n}}} \overset{a}{\sim} N(0,1)$$

<u>2. Quantis</u>



$b: P\{-b \leq T \leq b\} = 1-\alpha \iff$

$b: \Phi(b) = 1 - \frac{\alpha}{2}$, como $1-\alpha = 0.95$

$b = \Phi^{-1}(0.975) = 1.96$

INSTITUTO SUPERIOR TÉCNICO

3. ICA

$$-b \leq \frac{\bar{X} - p}{\sqrt{\frac{\bar{X}(1-\bar{X})}{n}}} \leq b \Leftrightarrow$$

$$\bar{X} - b\sqrt{\frac{\bar{X}(1-\bar{X})}{n}} \leq p \leq \bar{X} + b\sqrt{\frac{\bar{X}(1-\bar{X})}{n}}$$

$$ICA_{95\%}(p) \approx \left[\bar{X} - b\sqrt{\frac{\bar{X}(1-\bar{X})}{n}} \; ; \; \bar{X} + b\sqrt{\frac{\bar{X}(1-\bar{X})}{n}}\right]$$

4. IC

$$IC_{95\%}(p) \simeq \left[\bar{x} - b\sqrt{\frac{\bar{x}(1-\bar{x})}{n}} \; ; \; \bar{x} + b\sqrt{\frac{\bar{x}(1-\bar{x})}{n}}\right]$$

Recordando que: $b = 1.96$, $n = 250$, $\bar{x} = \frac{195}{250} = 0.78$

$$IC_{95\%}(p) \simeq [0.7286 \; ; \; 0.8314]$$

(b) Seja Δ_n a amplitude do IC determinado em (a)

$$\Delta_n = 2b\sqrt{\frac{\bar{x}(1-\bar{x})}{n}}$$

$$n: \; 2b\sqrt{\frac{\bar{x}(1-\bar{x})}{n}} = 0.04$$

$$\Leftrightarrow \left(\frac{0.02}{b}\right)^2 n = \bar{x}(1-\bar{x})$$

$$\Leftrightarrow \quad n = \frac{\bar{x}(1-\bar{x})b^2}{(0.02)^2}$$

Considerando que $1-\alpha = 0.95$ e por isso $b = 1.96$ e $\bar{x} \simeq \frac{195}{250} = 0.78$ é próximo do valor observado em (a) então $n = \lceil 1648.046 \rceil = 1649$

INSTITUTO SUPERIOR TÉCNICO

Grupo II

1.

tempo espera	O_i	$\hat{p}_i$	$\hat{E}_i = n\hat{p}_i$
$]0,2]$	40	0.4866	48.66
$]2,4]$	36	0.2498	24.98
$]4,+\infty[$	24	0.2636	26.36
	$n=100$		

Hipóteses: $H_0: X \sim Exp(\lambda)$ vs $H_1: X \not\sim Exp(\lambda)$

Estatística Teste: (admitindo H_0 verd) $T_0 = \sum_{i=1}^{k} \frac{(O_i - \hat{E}_i)^2}{\hat{E}_i} \underset{H_0}{\overset{a}{\sim}} \chi^2_{(k-\beta-1)}$

Valor observado da Estatística de Teste:

$$\hat{p}_1 = \hat{P}(X \leq 2 \mid H_0 \text{ verd}) = F_{Exp(1/3)}(2)$$

$$= 1 - e^{-2/3} \simeq 0.4866 \quad e \quad \hat{E}_1 = n\hat{p}_1 = 48.66$$

$$\hat{p}_2 = \hat{P}(2 < X \leq 4 \mid H_0 \text{ verd}) = F_{Exp(1/3)}(4) - F_{Exp(1/3)}(2)$$

$$= e^{-2/3} - e^{-4/3} \simeq 0.2498 \quad e \quad \hat{E}_2 = 24.98$$

$$\hat{p}_3 = \hat{P}(X > 4 \mid H_0 \text{ verd}) = 1 - F_{Exp(1/3)}(4) =$$

$$= e^{-4/3} \simeq 0.2636 \quad e \quad \hat{E}_3 = 26.36$$

Como $\hat{E}_i \geq 5$, $i=1,2,3$ então: não há necessidade de de agrupar classes e $k=3$, $\beta=1$.

Logo $t_0 = \frac{(40-48.66)^2}{48.66} + \frac{(36-24.98)^2}{24.98} + \frac{(24-26.36)^2}{26.36}$

$= 6.6140$

Valor-p:

$p = P(T_0 \geq t_0 \mid H_0 \text{ verd}) = 1 - F_{\chi^2_{(3-1-1)}}(6.6140)$

(i) Usando a máquina de calcular:

$p \cong 1 - 0.9899 = 0.0101$

(ii) Usando as tabelas, sabe-se que:

$F_{\chi^2_{(1)}}(5.024) = 0.975$

$F_{\chi^2_{(1)}}(6.635) = 0.990$

Pode então escrever-se:

$F_{\chi^2_{(1)}}(5.024) \leq F_{\chi^2_{(1)}}(6.6140) \leq F_{\chi^2_{(1)}}(6.635)$

$1 - 0.990 \leq 1 - F_{\chi^2_{(1)}}(6.6140) \leq 1 - 0.975$

$0.01 \leq p \leq 0.025$

Regra de Decisão:

(i) Máquina calcular:

$\alpha \geq 0.0101 \Rightarrow$ Rejeito H_0

$\alpha < 0.0101 \Rightarrow$ não Rejeito H_0

(ii) Tabelas

$\alpha \geq 0.025 \Rightarrow$ Rej H_0

$\alpha \leq 0.01 \Rightarrow \bar{n}$ Rej H_0

$0.025 < \alpha < 0.01 \Rightarrow$ inconclusivo

INSTITUTO SUPERIOR TÉCNICO

Decisão:

Aos níveis de significância , 5% e 10% Rejeito H_0

2. (a)

$$\hat{\beta}_1 = \frac{\sum x_i y_i - n\bar{x}\bar{y}}{\sum x_i^2 - n\bar{x}^2} =$$

$$= \frac{6542.9 - 271.7 * 263.8/9}{9258.69 - (271.7)^2/9}$$

$$= \frac{-1420.929}{1056.369} \simeq -1.345107$$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} =$$

$$= 69.91838$$

Logo $\hat{y}_i = 69.9184 - 1.3451 x_i$

Este modelo estimado, após validade, só deve ser usado quando

$$16.0 \leq x_i \leq 43.9$$

2 (b) Hipóteses: $H_0: \beta_1 = 0$ vs $H_1: \beta_1 \neq 0$

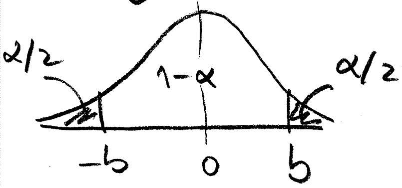
Variável fulcral: $T = \dfrac{\hat{\beta}_1 - \beta_1}{se(\hat{\beta}_1)} \sim t_{(n-2)}$

onde $se(\hat{\beta}_1) = \sqrt{\dfrac{\hat{\sigma}^2}{\sum x_i^2 - n\bar{x}^2}}$

Estatística Teste: $T_0 = T \mid H_0$ verd

$$= \frac{\hat{\beta}_1}{se(\hat{\beta}_1)} \sim t_{(n-2)}$$

Região Crítica: Sendo $\alpha = 0.10$



b: $P(T_0 \leq b \mid H_0 \text{ verd}) = 1 - \frac{\alpha}{2}$

$$b = F^{-1}_{t_{(9-2)}}(0.95)$$

$$b = 1.895$$

$RC =]-\infty, -1.895[\cup]1.895, +\infty[$

Regra de Decisão:

Se $|t_0| \geq 1.895$ então Rejeito H_0

Se $|t_0| < 1.895$ então não Rejeito H_0

Decisão:

$$\hat{\sigma}^2 = \frac{1}{n-2}\left\{ \sum y_i^2 - n\bar{y}^2 - (\hat{\beta}_1)^2 \left[\sum x_i^2 - n\bar{x}^2\right] \right\}$$

$$= 58.79547$$

$t_0 = -5.7015 < -1.895$ logo Rejeito H_0 ao nível de significância de 10%.

Assim, tendo por base esta amostra, há evidência para acreditar que a taxa de natalidade explica de forma linear a taxa de urbanização.