2º TESTE (9 JUN 2011, código C)

1. Seja X a v.a. que indica o tempo (em segundos) entre chegadas consecutivas de chamadas.

$$X \sim Exp(\lambda) \iff f_X(x) = \begin{cases} \lambda e^{-\lambda x} & , x \geq 0 \\ 0 & , x < 0 \end{cases} \qquad \lambda > 0$$

$\underset{\sim}{X} = (X_1, \dots, X_{100})$ a.a. de dimensão $n = 100$ da pop. X

$\underset{\sim}{x} = (x_1, \dots, x_{100}) \rightarrow \sum_{i=1}^{100} x_i = 1\ 121$ segundos

a) Método da Máxima Verosimilhança (MMV)

- A função de verosimilhança duma amostra é

$$L(\lambda | \underset{\sim}{x}) = f_{X_1,\dots,X_n}(x_1,\dots,x_n) \underset{X_i \text{ ind.}}{=} \prod_{i=1}^{n} f_{X_i}(x_i) \underset{X_i \overset{d}{=} X}{=} \prod_{i=1}^{n} \lambda e^{-\lambda x_i} =$$

$$= \lambda^n e^{-\lambda \sum_{i=1}^{n} x_i} \quad , \lambda > 0$$

- Maximizar $L \iff$ maximizar $\ln L$

$$\ln L = n \ln \lambda - \lambda \sum_{i=1}^{n} x_i$$

$$\left.\frac{d \ln L}{d\lambda}\right|_{\lambda = \hat{\lambda}} = 0 \Rightarrow \frac{n}{\hat{\lambda}} - \sum_i x_i = 0 \Rightarrow \hat{\lambda} = \frac{n}{\sum_i x_i} = \frac{1}{\bar{x}}$$

$$\frac{d^2 \ln L}{d\lambda^2} = -\frac{n}{\lambda^2} < 0 \ , \forall_\lambda$$

Assim, a estimativa de M.V. de λ é $\hat{\lambda} = \frac{100}{1121} = 0.0892$

b) $P(X<5) = \int_0^5 \lambda e^{-\lambda x} dx = \left[-e^{-\lambda x}\right]_0^5 = 1 - e^{-5\lambda} = g(\lambda)$

Pela propriedade da Invariância dos estimadores de MV,

$$\hat{P}(X<5) = g(\hat{\lambda}) = 1 - e^{-5\hat{\lambda}} = 1 - e^{-5 \times 0.0892} = 0.3598$$

2. Seja $Y_i \to$ concentração de partículas em suspensão no ar (microgramas por m³) na proximidade da cimenteira i, $i=1,2$.

$$Y_i \sim N(\mu_i, \sigma_i^2) \quad \text{com} \quad \sigma_1^2 = \sigma_2^2 = \sigma^2$$

Das amostras observadas, obtém-se:

$$\begin{cases} n_1 = 10 \\ \bar{y}_1 = \frac{850}{10} = 85 \\ (n_1-1)s_1^2 = \sum_i y_{1i}^2 - n_1 \bar{y}_1^2 = 1000 \end{cases} \qquad \begin{cases} n_2 = 15 \\ \bar{y}_2 = \frac{1350}{15} = 90 \\ (n_2-1)s_2^2 = \sum_i y_{2i}^2 - n_2 \bar{y}_2^2 = 1500 \end{cases}$$

a) Método da variável fulcral

- variável fulcral (Pop. normais com variâncias iguais mas desconhecidas)

$$T = \frac{\bar{Y}_1 - \bar{Y}_2 - (\mu_1-\mu_2)}{\sqrt{\frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1+n_2-2}\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}} \sim t_{(n_1+n_2-2)} \equiv t_{23}$$

- Determinação dos quantis (simétricos), $\pm a$:

$$a = F^{-1}_{t_{23}}(1-\alpha/2) = F^{-1}_{t_{23}}(0.99) = 2.5$$

$$P\left(-2.5 \leq \frac{\bar{Y}_1-\bar{Y}_2-(\mu_1-\mu_2)}{\sqrt{\quad}} \leq 2.5\right) = 0.98$$

- Determinação do Intervalo de Confiança Aleatório

$$P\left(\bar{Y}_1-\bar{Y}_2 - 2.5\sqrt{\quad} \leq \mu_1-\mu_2 \leq \bar{Y}_1-\bar{Y}_2 + 2.5\sqrt{\quad}\right) = 0.98$$

$$ICA_{98\%}(\mu_1-\mu_2) = \left(\bar{Y}_1-\bar{Y}_2 \pm 2.5\sqrt{\quad}\right)$$

- concretização

$$IC_{98\%}(\mu_1-\mu_2) = \left(85-90 \pm 2.5\underbrace{\sqrt{\frac{1000+1500}{23}\left(\frac{1}{10}+\frac{1}{15}\right)}}_{18.116}\right) = (-15.64, 5.64)$$

b) Pretende-se testar, para $\alpha = 2\%$, a hipótese $H_0: \mu_1-\mu_2=0$ versus $H_1: \mu_1-\mu_2 \neq 0$.

Invocando a relação entre intervalos de confiança e testes de hipóteses, pode-se efectuar a decisão com base no intervalo obtido, da seguinte forma:

- Como $\mu_1-\mu_2 = 0 \in IC_{98\%}(\mu_1-\mu_2)$, não é de rejeitar a hipótese de $\mu_1=\mu_2$, para o nível de significância 2%.

Grupo II

1. Seja $X \to$ nº mensal de acidentes num dado cruzamento

Amostra de 120 meses, agrupada, está na tabela:

x_i	o_i	$E_i = 120\, p_i^0$	$(o_i - E_i)^2/E_i$
0	79	72.78368	0.5309
1	27	36.39184	2.4238
2 (≥ 2)	12 (14)	9.09796 (10.824481)	0.9316
≥ 3 (≥ 2)	2 (14)	1.726521 (10.824481)	
total	120	120	$3.8863 = q_0$

Teste de Ajustamento

- H_0: $X \sim Poisson(0.5)$ vs H_1: $X \not\sim Poisson(0.5)$
- Estatística de teste (Estatística de χ^2 de Pearson)

$$Q_0 = \sum_{i=1}^{K} \frac{(O_i - E_i)^2}{E_i} \underset{H_0}{\overset{a}{\sim}} \chi^2_{(K-\beta-1)}$$

$$E_i = E(O_i \mid \text{sob } H_0) = n\, p_i^0 \quad , i = 1, \dots, K$$

$$p_i^0 = P(X \in \text{classe } i \mid X \sim Poi(0.5))$$

Sob validade de H_0, $P(X=x) = \frac{0.5^x}{x!} e^{-0.5}$, $x = 0, 1, 2, \dots$

$$p_1^0 = P(X=0) = e^{-0.5} = 0.6065307$$

$$p_2^0 = P(X=1) = 0.5\, e^{-0.5} = 0.3032653$$

$$p_3^0 = P(X=2) = \frac{0.5^2}{2} e^{-0.5} = 0.07581633$$

$$p_4^0 = P(X \geq 3) = 1 - \sum_{i=1}^{3} p_i^0 = 0.001751623$$

- Cálculo do valor-p: Sendo $K = 3$ e $\beta = 0$

$$p = P(Q > q_0 \mid \text{sob } H_0) = P(Q_0 > 3.8863) = 1 - F_{\chi^2_2}(3.8863) = 0.143252$$

- Decisão com o valor-p:
  - Rejeita-se H_0 para $\alpha > 14.3252\%$
  - Não se Rejeita H_0 para $\alpha \leq 14.3252\%$, pelo que as observações são consistentes com $X \sim Poi(0.5)$ para níveis usuais de α.

Alternativa (usando tabela):

$$0.85 < F_{\chi^2_2}(3.8863) < 0.90 \Rightarrow 0.10 < p < 0.15$$

- Rej. H_0 para $\alpha \geq 15\%$
- Não Rej. H_0 para $\alpha \leq 10\%$, pelo que ...

2. Seja $X \to$ altura (em cm) de indivíduos
e $Y \to$ nº pulsações por minuto após actividade física

Modelo de RLS de Y em X $\begin{cases} Y = \beta_0 + \beta_1 x + \varepsilon \\ E(Y|x) = \beta_0 + \beta_1 x \end{cases}$, com $E(\varepsilon)=0$, $V(\varepsilon)=\sigma^2$

Pelo MMQ, as estimativas de β_0 e β_1 são:

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} \simeq 14.819$$

$$\hat{\beta}_1 = \frac{\sum x_i y_i - m\bar{x}\bar{y}}{\sum x_i^2 - m\bar{x}^2} = 0.70207 \simeq 0.702$$

Cálculos Auxiliares:

$$\sum x_i^2 - m\bar{x}^2 = 336\,752 - 12 \times 165.5^2 = 8069$$

$$\sum y_i^2 - m\bar{y}^2 = 225\,932 - 12 \times 131^2 = 20\,000 = 2\times 10^4$$

$$\sum x_i y_i - m\bar{x}\bar{y} = \hat{\beta}_1 (\sum x_i^2 - m\bar{x}^2) = 5665$$

a) Teste sobre β_1

Se $\beta_1 = 0 \Rightarrow E(Y|x) = \beta_0 + \beta_1 x = \beta_0$ (modelo não é adequado, a regressão não é significativa)

- $H_0: \beta_1 = 0$ vs $H_1: \beta_1 \neq 0$
- Estatística de teste (T_0)

$$T = \frac{\hat{\beta}_1 - \beta_1}{\sqrt{\frac{\hat{\sigma}^2}{\sum x_i^2 - m\bar{x}^2}}} \sim t_{(m-2)} \to T_0 = \frac{\hat{\beta}_1 - 0}{\sqrt{-}} \underset{H_0}{\sim} t_{10}$$

$$t_0 = \frac{0.70207 - 0}{\sqrt{\frac{1602.278}{8069}}} = 1.5755$$

onde $\hat{\sigma}^2 = \frac{1}{m-2}\left[\sum y_i^2 - m\bar{y}^2 - \hat{\beta}_1^2(\sum x_i^2 - m\bar{x}^2)\right] = \frac{2\times 10^4 - 0.70207^2 \times 8069}{10}$
$= 1602.278$

- Região crítica de nível α: $R_\alpha = \{ |T_0| > a = F^{-1}_{t_{10}}(1-\alpha/2) \}$
  Para $\alpha = 1\%$, $R_{0.01} = \{ |T_0| > 3.169 \}$
- Decisão: Como $t_0 = 1.5755 \notin R_{0.01}$, não se rejeita H_0 pelo que a hipótese de a recta de regressão não ser significativa não é rejeitada para o nível $\alpha = 1\%$.

b) Coeficiente de determinação

$$R^2 = \frac{(\Sigma u_i Y_i - m\bar{u}\bar{Y})^2}{(\Sigma u_i^2 - m\bar{u}^2)(\Sigma Y_i^2 - m\bar{Y}^2)} = \hat{\beta}_1^2 \times \frac{\Sigma u_i^2 - m\bar{u}^2}{\Sigma Y_i^2 - m\bar{Y}^2}$$

$$r^2 = 0.70207^2 \times \frac{8069}{2\times 10^4} = 0.1988 \simeq 0.20$$

Apenas cerca de 20% da variação total do nº de pulsações por minuto, após a actividade física, é explicada pela altura dos indivíduos, pelo que a recta estimada é um mau ajustamento.

Este resultado está em consonância com o da alínea anterior, uma vez que a hipótese de a regressão não ser significativa não foi rejeitada.