2º teste (9/6/2011, código D)

Grupo I

1. Seja $X \to$ tempo (em minutos) entre chegadas consecutivas de veículos a uma estação de serviço

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

$\underset{\sim}{X} = (X_1, \dots, X_{200})$ a.a. de dimensão $n = 200$ da pop. X

$\underset{\sim}{x} = (x_1, \dots, x_{200}) \to \sum_{i=1}^{200} x_i = 2154$ minutos

a) Método da Máxima verosimilhança (MMV)

• A função de verosimilhança duma amostra $\underset{\sim}{x} = (x_1, \dots, x_n)$ é

$$L(\lambda \mid \underset{\sim}{x}) = f_{X_1,\dots,X_n}(x_1,\dots,x_n) \underset{X_i \text{ ind.}}{=} \prod_{i=1}^{n} f_{X_i}(x_i) \underset{X_i \overset{d}{=} X}{=} \prod_{i=1}^{n} \lambda e^{-\lambda x_i} = \lambda^n e^{-\lambda \sum x_i}, \lambda > 0$$

• Maximizar $L \Leftrightarrow$ Maximizar $\ln L$

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

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

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

∴ A estimativa M.V. de λ é $\hat{\lambda} = \frac{200}{2154} = 0.09285$ //

b) $P(X > 10) = \int_{10}^{+\infty} \lambda e^{-\lambda x} dx = \left[-e^{-\lambda x}\right]_{10}^{+\infty} = e^{-10\lambda} = g(\lambda)$

Pela propriedade da Invariância dos Estimadores de M.V.

$\hat{P}(X > 10) = g(\hat{\lambda}) = e^{-10\hat{\lambda}} = e^{-10 \times 0.09285} = 0.39515$ //

2. Seja $X_i \rightarrow$ concentração de partículas em suspensão no ar (em microgramas por m^2) na cidade i, $i=1,2$.

$X_i \sim N(\mu_i, \sigma_i^2)$ com $\sigma_1^2 = \sigma_2^2 = \sigma^2$

Das amostras observadas, obtém-se:

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

a) Método da variável fulcral

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

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

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

$$a = F^{-1}_{t_{(n_1+n_2-2)}}(1-\alpha/2) = F^{-1}_{t_{23}}(0.995) = 2.807$$

$$P\left(-2.807 \leq \frac{\bar{X}_1-\bar{X}_2-(\mu_1-\mu_2)}{\sqrt{\quad}} \leq 2.807\right) = 0.99$$

- Determinação do Intervalo de Confiança Aleatório

$$P\left(\bar{X}_1-\bar{X}_2 - 2.807\sqrt{\;-\;} \leq \mu_1-\mu_2 \leq \bar{X}_1-\bar{X}_2 + 2.807\sqrt{\;-\;}\right) = 0.99$$

$$ICA_{99\%}(\mu_1-\mu_2) = \left(\bar{X}_1-\bar{X}_2 \pm 2.807\sqrt{\;-\;}\right)$$

- concretização

$$IC_{99\%}(\mu_1-\mu_2) = \left(90-85 \pm 2.807 \underbrace{\sqrt{\frac{1500+1000}{23}\left(\frac{1}{15}+\frac{1}{10}\right)}}_{18.116}\right) = (5 \pm 11.947) =$$
$$= (-6.947\,,\,16.947)$$

b) Pretende-se testar, para $\alpha = 1\%$, 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 $\mu_1-\mu_2 = 0 \in IC_{99\%}(\mu_1-\mu_2)$, não é de rejeitar a hipótese de $\mu_1=\mu_2$, para o nível de significância 1%.

Grupo II

1. Seja $X \to$ nº de defeitos em certo tipo de circuitos

Amostra de $n=100$ circuitos, agrupada, está na tabela:

u_i	o_i	$E_i = 100\, p_i^0$	$(o_i - E_i)^2/E_i$
0	52	44.93	1.1125
1	22	35.95	5.413
≥ 2: 2	19 } 26	14.38 } 19.12	2.4756
≥ 2: ≥ 3	7	4.74	
total	100	100	$9.001 = q_0$

Teste de Ajustamento

- $H_0: X \sim \text{Poisson}(0.8)$ vs $H_1: X \not\sim \text{Poi}(0.8)$
- Estatística de teste (Estatística de χ^2 de Pearson)

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

$$E_i = E(O_i \mid \text{sob } H_0) = n\, p_i^0 \quad , \quad p_i^0 = p(X \in \text{classe } i \mid X \sim \text{Poi}(0.8))$$

Sob validade de H_0, $p(X=u) = \frac{0.8^u}{u!} e^{-0.8}$, $u = 0, 1, 2, \ldots$

$p_1^0 = p(X=0) = e^{-0.8} = 0.4493$

$p_2^0 = p(X=1) = 0.8\, e^{-0.8} = 0.3595$

$p_3^0 = p(X=2) = \frac{0.8^2}{2} e^{-0.8} = 0.1438$

$p_4^0 = p(X \geq 3) = 1 - \sum_{i=1}^{3} p_i^0 = 1 - 0.9526 = 0.0474$

- Região crítica de nível α, $R_\alpha = \{ Q_0 > a = F^{-1}_{\chi^2_{(K-\beta-1)}}(1-\alpha) \}$

$K = 3$ e $\beta = 0$

Cálculo do valor-p:

$$p = P(Q > q_0 \mid \text{sob } H_0) = P(Q_0 > 9.001) = 1 - F_{\chi^2_2}(9.001) \simeq 0.0111$$

Usando tabela: $0.975 < F_{\chi^2_2}(9.001) < 0.99 \Rightarrow 0.01 < p < 0.025$

- Decisão com base no intervalo obtido para o valor-p:
  - Rej. H_0 para $\alpha \geq 2.5\%$
  - Não Rej. H_0 para $\alpha \leq 1\%$
  - Entre 1% e 2.5% nada se poderá decidir.

Decisão com o valor-p:
- Não Rejeitar H_0 para $\alpha \leq 1.11\%$
- Rejeitar H_0 para $\alpha > 1.11\%$ (A evidência contra H_0 é elevada)

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

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:

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} \simeq 59.208$$

$$\hat{\beta}_1 = \frac{\sum x_i y_i - n\bar{x}\bar{y}}{\sum x_i^2 - n\bar{x}^2} = 0.139975 \simeq 0.14$$

Cálculos Auxiliares:

$$\sum_i x_i^2 - n\bar{x}^2 = 1\,347\,006 - 47 \times \left(\frac{7946}{47}\right)^2 = 3\,624.808512$$

$$\sum_i y_i^2 - n\bar{y}^2 = 326\,241 - 47 \times \left(\frac{3895}{47}\right)^2 = 3\,453.234043$$

$$\sum_i x_i y_i - n\bar{x}\bar{y} = \hat{\beta}_1 \left(\sum x_i^2 - n\bar{x}^2\right) = 507.3829787$$

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)

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

$$t_0 = \frac{0.14}{\sqrt{\frac{75.1603}{3624.808512}}} = \frac{0.14}{0.143996} = 0.9723$$

onde $\hat{\sigma}^2 = \frac{1}{n-2}\left[\sum y_i^2 - n\bar{y}^2 - \hat{\beta}_1^2 \left(\sum x_i^2 - n\bar{x}^2\right)\right] = \frac{3382.21305}{45} = 75.1603$

- Região crítica de nível α: $R_\alpha = \{ |T_0| > a = F^{-1}_{t_{45}}(1-\alpha/2) \}$

Para $\alpha = 5\%$, $R_{0.05} = \{ |T_0| > 2.014 \}$

- Decisão: Como $t_0 = 0.9723 \notin R_{0.05}$, não se rejeita H_0, pelo que a hipótese de a recta de regressão não ser significativa não é rejeitada para o nível $\alpha = 5\%$ (e também para $\alpha \leq 5\%$).

b) Coeficiente de determinação

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

$$r^2 = 0.14^2 \times \frac{3624.808512}{3453.239043} = 0.0205$$

Apenas 2.05 % da variação total do nº de pulsações por minuto em repouso dos indivíduos é explicada pela altura pelo que a recta estimada é um ajustamento muito mau.

Este resultado corrobora o da alínea anterior, uma vez que a hipótese de a recta de regressão não ser significativa foi rejeitada.