Departamento de Matemática
Probabilidades e Estatística
MEFT + MEBM + MEQ + LEMat

Segundo teste - versão 2 — 2º semestre – 2010/2011
Duração: 90 minutos — 09/06/2011 – 16:00

▼ Justifique convenientemente todas as respostas ▼

Grupo I — 10 valores

1. Admita que a variável aleatória X, que indica o tempo (em minutos) entre chegadas consecutivas de veículos a uma estação de serviço, possui distribuição exponencial de parâmetro λ. Uma concretização de uma amostra aleatória de dimensão 200 da variável X conduziu a $\sum_{i=1}^{200} x_i = 2154$ minutos.

(a) Determine a estimativa de máxima verosimilhança de λ com base na amostra dada. (3.0)

Função de verosimilhança: $\mathcal{L}(\lambda; x_1, \ldots, x_n) \overset{i.i.d.}{=} \prod_{i=1}^n f_X(x_i) = \prod_{i=1}^n \lambda\, e^{-\lambda x_i} = \lambda^n e^{-\lambda \sum_{i=1}^n x_i}$
$\frac{d \log \mathcal{L}}{d\lambda} = 0 \quad \Leftrightarrow \quad \frac{n}{\lambda} - \sum_{i=1}^n x_i = 0 \quad \Leftrightarrow \quad \lambda = \bar{x}^{-1}$ e $\frac{d^2 \log \mathcal{L}}{d\lambda^2} = -\frac{n}{\lambda^2} < 0,\ \forall \lambda > 0, \forall n \in I\!N.$
$\therefore$ Estimador de máxima verosimilhança (MV): $\hat{\lambda} = \bar{X}^{-1}$, cuja estimativa é $(\frac{2154}{200})^{-1} \approx 0.0928$.

(b) Obtenha, para a amostra dada, a estimativa de máxima verosimilhança da probabilidade do tempo entre chegadas consecutivas de veículos ser superior a 10 minutos. (2.0)

Sendo $\phi = P(X > 10) = \int_{10}^{\infty} \lambda\, e^{-\lambda x} dx = e^{-10\lambda}$, o seu estimador de MV é $\hat{\phi} = e^{-10\bar{X}^{-1}}$, pela propriedade de invariância dos estimadores de MV.
Na amostra observada, a estimativa de MV de ϕ é $e^{-2000/2154} \approx 0.3651$.

2. Para avaliar a qualidade do ar de duas cidades, A e B, consideraram-se as variáveis aleatórias X_1 e X_2, que indicam a concentração de partículas em suspensão no ar (em microgramas por m^3) nas cidades A e B, respectivamente. Suponha que X_1 e X_2 têm distribuições normais com variâncias desconhecidas mas iguais. A concretização de duas amostras aleatórias independentes, de dimensões 15 e 10, de X_1 e X_2, conduziu aos seguintes valores:

$$\sum_{i=1}^{15} x_{1i} = 1350; \qquad \sum_{i=1}^{10} x_{2i} = 850; \qquad \sum_{i=1}^{15} x_{1i}^2 = 123000; \qquad \sum_{i=1}^{10} x_{2i}^2 = 73250.$$

(a) Determine um intervalo de confiança, com nível de confiança de 99%, para a diferença entre os valores esperados das concentrações de partículas em suspensão no ar nas cidades A e B. (3.5)

Se $Y_1 \sim N(\mu_1, \sigma^2)$ e $Y_2 \sim N(\mu_2, \sigma^2)$, independentes, a variável fulcral do intervalo de confiança (IC) de $\mu_1 - \mu_2$ é $W = \frac{\bar{X}_1 - \bar{X}_2 - (\mu_1 - \mu_2)}{\sqrt{\frac{14\, S_1^2 + 9\, S_2^2}{23}\left(\frac{1}{15} + \frac{1}{10}\right)}} \sim t_{(23)}$. Assim, $P(-2.807 < W < 2.807) = 0.99 \Leftrightarrow$
$P\left(\bar{X}_1 - \bar{X}_2 - 2.807\sqrt{\frac{14S_1^2 + 9S_2^2}{23}\left(\frac{1}{15} + \frac{1}{10}\right)} < \mu_1 - \mu_2 < \bar{X}_1 - \bar{X}_2 + 2.807\sqrt{\frac{14S_1^2 + 9S_2^2}{23}\left(\frac{1}{15} + \frac{1}{10}\right)}\right) = 0.99$.
$\therefore IAC(\mu_1 - \mu_2; 0.99) = \left(\bar{X}_1 - \bar{X}_2 \pm 2.807\sqrt{\frac{14\, S_1^2 + 9\, S_2^2}{23}\left(\frac{1}{15} + \frac{1}{10}\right)}\right)$, cuja concretização é dada por $IC(\mu_1 - \mu_2; 0.99) = 85 - 90 \pm 2.807\sqrt{\frac{14 \times 107.143 + 9 \times 111.111}{23}\left(\frac{1}{15} + \frac{1}{10}\right)} = (-6.96, 16.96)$.

(b) Teste, ao nível de significância de 1%, a conjectura da igualdade dos valores esperados das concentrações de partículas em suspensão no ar nas cidades A e B, com base no intervalo de confiança obtido na alínea anterior. (1.5)

Pretende-se testar $H_0 : \mu_1 = \mu_2$ contra $H_1 : \mu_1 \neq \mu_2$, usando $\alpha = 0.01$, com base no $IC(\mu_1 - \mu_2; 0.99)$ em a). Como $0 \in (-6.96, 16.96)$, não há nada contra a conjectura da igualdade dos valores esperados das concentrações de partículas em suspensão no ar nas cidades A e B (H_0) para um nível de 1% de significância, visto que com um grau de 99% de confiança, $\mu_1 - \mu_2 = 0$ é um valor plausível para essa diferença de valores esperados.

Grupo II	10 valores

1. Pensa-se que a variável aleatória X, que denota o número de defeitos encontrados em certo tipo de circuitos eléctricos, tem distribuição de Poisson com valor esperado 0.8. Tendo-se inspeccionado 100 circuitos do tipo referido, escolhidos ao acaso, os números de defeitos neles encontrados foram agrupados na seguinte tabela: Será que os dados corroboram a hipótese de X seguir a distribuição

nº de defeitos	0	1	2	3 ou mais
frequência	52	22	19	7

indicada? Recorra para o efeito ao cálculo do valor–p. (4.0)

Seja X o número mensal de acidentes nesse cruzamento. Considerem-se inicialmente as classes $C_1 = \{0\}$, $C_2 = \{1\}$, $C_3 = \{2\}$ e $C_4 = \{3, \ldots\}$, fazendo junção de classes quando não houver i) todos E_i maiores ou iguais a 1 e ii) pelo menos 80% deles maiores ou iguais a 5, onde $p_i = P(X \in C_i)$, $i = 1, 2, 3, 4$. Pretende-se testar a hipótese $H_0 : X \sim \text{Poisson}(\lambda = 0.8)$ contra $H_1 : X \nsim \text{Poisson}(\lambda = 0.8)$, recorrendo à estatística de Pearson

$$Q = \sum_{i=1}^{3} \frac{(O_i - E_i)^2}{E_i} \underset{H_0}{\overset{a}{\sim}} \chi^2_{(2)}.$$

Junção de classes: Todos E_i maiores ou iguais a 1 e pelo menos 80% deles devem ser no mínimo 5.

Classes	O_i	$E_i = np_i^0$	$(O_i - E_i)^2/E_i$
C_1	52	$100 \times \mathrm{e}^{-0.8} = 44.933$	1.111
C_2	22	$100 \times \mathrm{e}^{-0.8} \times 0.8 = 35.946$	5.411
C_3	19 ⌉ 26	$100 \times \mathrm{e}^{-0.8} \times 0.8^2/2 = 14.378$ ⌉ 19.120	2.476
(C_4)	7 ⌋	$100 \times (1 - \sum_{x=0}^{2} \frac{\mathrm{e}^{-0.8} 0.8^x}{x!}) = 4.742$ ⌋	
	$n = 100$		$q = 8.998$

Valor-$p = P(Q \geq q | H_0) = 1 - F_{\chi^2_{(2)}}(8.998) \approx 1 - 0.9899 = 0.0111$.
A hipótese H_0 não deve ser rejeitada para $\alpha > 0.1431$ e deve ser rejeitada no caso contrário. Como o valor-p é superior aos valores usuais para o nível de significância de um teste de hipóteses (1%, 5% e 10%), há evidência a favor da hipótese de o número mensal de acidentes nesse cruzamento seguir a distribuição de Poisson de parâmetro 0.5.

2. Com o objectivo de estudar, em seres humanos, a relação entre a altura (x, em cm) e o número de pulsações por minuto em repouso, Y, fizeram-se medições em 47 pessoas, tendo-se obtido os seguintes resultados:

$$\sum_{i=1}^{47} x_i = 7946 \quad \sum_{i=1}^{47} x_i^2 = 1347006 \quad \sum_{i=1}^{47} y_i = 3895 \quad \sum_{i=1}^{47} y_i^2 = 326241 \quad \sum_{i=1}^{47} x_i y_i = 659011$$

As medições efectuadas conduziram à seguinte estimativa da recta de mínimos quadrados: $\widehat{E(Y|x)} = 59.208 + 0.140\,x$. Tendo em conta o modelo de regressão linear simples $Y_i = \beta_0 + \beta_1 x_i + \epsilon_i$, $i = 1, \ldots, 47$, com as hipóteses de trabalho habituais, responda às questões seguintes:

(a) Teste, ao nível de significância de 5%, a significância da recta de regressão. (4.0)

Pretende-se testar as hipóteses $H_0 : \beta_1 = 0$ contra $H_1 : \beta_1 \neq 0$ com base na estatística do teste $T = \frac{\hat{\beta}_1}{\sqrt{\frac{\hat{\sigma}^2}{\sum x_i^2 - n\bar{x}^2}}} \overset{H_0}{\sim} t_{(45)}$, cujo valor observado é $t = \frac{0.14}{\sqrt{\frac{75.1667}{3624.808}}} \approx 0.972$. Note-se que $\hat{\sigma}^2 = \frac{1}{45}[326241 - 47 \times 82.872^2 - 0.14^2(1347006 - 47 \times 169.064^2)] \approx 75.1667$. Para um nível de significância de 5%, a região crítica é $RC_{5\%} = (-\infty, -2.014) \cup (2.014, \infty)$ pois $F^{-1}_{t_{(45)}}(0.975) = 2.014$. Como $t \notin RC_{5\%}$, não se rejeita a hipótese H_0 ao nível de significância de 5%, ou seja, a recta de regressão não é significativa i.e. a altura não influencia o número de pulsações por minuto logo após a realização de uma determinada actividade física, a esse nível de significância.

(b) Obtenha o coeficiente de determinação do modelo. Comente os resultados obtidos, tendo em conta não só o valor deste coeficiente mas também o resultado da alínea anterior. (2.0)

$$R^2 = \frac{\left(\sum_{i=1}^{n} x_i Y_i - n\bar{x}\bar{Y}\right)^2}{\left(\sum_{i=1}^{n} x_i^2 - n\bar{x}^2\right) \times \left(\sum_{i=1}^{n} Y_i^2 - n\bar{Y}^2\right)} = \frac{(507.820)^2}{3653.549 \times 3624.808} \approx 0.0195.$$

Cerca de 2% da variação total do número de pulsações por minuto após a actividade física é explicada pelo modelo de regressão linear simples com a altura como variável independente, indicando assim um mau ajustamento desse modelo.

Esse mau ajustamento já era previsível com a aceitação da hipótese de não significância da recta de regressão (hipótese H_0) na alínea anterior.