Departamento de Matemática

# Probabilidades e Estatística

**MEC + LEGM + LET + LEIC-A**

Segundo teste (recurso) — 1º semestre – 2010/2011

Duração: 90 minutos — 03/02/2011 – 11:30

**▼ Justifique convenientemente todas as respostas ▼**

| **Grupo I** | 5 valores |
|---|---|

**1.** Seja $(X_1, \ldots, X_n)$ uma amostra aleatória de dimensão $n$ de uma população $X$ em que $E(X) = 3\theta$ e $Var(X) = 3\theta^2$, sendo $\theta \in I\!R$ um parâmetro cujo valor se desconhece.

(a) Averigue se os estimadores $T_1 = \frac{\bar{X}}{3}$ e $T_2 = \frac{n\bar{X}+1}{3n}$ são ambos não enviesados para $\theta$, $\forall n \in I\!N$. (1.0)

$E[T_1] = \frac{E[\bar{X}]}{3} = \frac{E[X]}{3} = \theta$ e $E[T_2] = \frac{E[\bar{X}]}{3} + \frac{1}{3n} = \frac{E[X]}{3} + \frac{1}{3n} = \theta + \frac{1}{3n} \neq \theta$, $\forall n \in I\!N$.
$\therefore T_1$ é centrado para $\theta$ ao contrário de $T_2$ que só é assintoticamente centrado.

(b) Qual dos estimadores $T_1$ ou $T_2$ é mais eficiente na estimação de $\theta$ para uma dimensão amostral finita? (1.0)

$Var[T_1] = Var[T_2] = \frac{Var[\bar{X}]}{9} = \frac{Var[X]}{9n} = \frac{\theta^2}{3n}$.
$e(T_1, T_2) = \frac{EQM[T_1]}{EQM[T_2]} = \frac{Var[T_1]+(E[T_1]-\theta)^2}{Var[T_2]+(E[T_2]-\theta)^2} = \frac{\frac{\theta^2}{3n}}{\frac{\theta^2}{3n}+\frac{1}{9n^2}} < 1$, $\forall n \in I\!N$, $\forall \theta \in I\!R$.
$\therefore T_1$ é um estimador de $\theta$ mais eficiente que $T_2$.

**2.** Admita que o número de erros de impressão por página de um livro é uma variável aleatória $X$ com distribuição de Poisson de parâmetro $\lambda \in I\!R^+$.

(a) Com base em uma amostra aleatória $(X_1, \ldots, X_n)$ de $X$ obtenha o estimador de máxima verosimilhança da função paramétrica definida por $E(X^2)$. (1.5)

Pretende-se estimar $\psi(\lambda) = E[X^2] = Var[X] + E^2[X] = \lambda + \lambda^2 = \lambda(1+\lambda)$.

$\mathcal{L}(\lambda; x_1, \ldots, x_n) = \prod_{i=1}^{n} \frac{e^{-\lambda}\lambda^{x_i}}{x_i!} = \frac{e^{-n\lambda}\lambda^{\sum_{i=1}^n x_i}}{\prod_{i=1}^n x_i!}$ e $\log\mathcal{L} = -n\lambda + \sum_{i=1}^n x_i \log\lambda - \log(\prod_{i=1}^n x_i!)$

$\frac{d\log\mathcal{L}}{d\lambda} = 0 \Leftrightarrow -n + \frac{\sum_{i=1}^n x_i}{\lambda} = 0 \Leftrightarrow -n\lambda + \sum_{i=1}^n x_i = 0 \Leftrightarrow \lambda = \frac{\sum_{i=1}^n x_i}{n} = \bar{x}$.

$\frac{d^2\log\mathcal{L}}{d\lambda^2} = -\frac{\sum_{i=1}^n x_i}{\lambda^2} < 0$, $\forall \lambda \in I\!R^+$ se $\exists i \in \{1, \ldots, n\} : x_i > 0$.

$\therefore \bar{X}$ é o estimador de máxima verosimilhança (EMV) de $\lambda$.

Pela invariância dos EMV's temos que $\hat{\psi}_{MV}(\lambda) = \psi(\hat{\lambda}_{MV}) = \bar{X}(1+\bar{X})$.

(b) A contagem dos erros de impressão em 100 páginas de um livro seleccionadas ao acaso produziu uma amostra na qual 25 páginas apresentam um erro, 8 páginas evidenciam 2 erros, 2 páginas contêm 3 erros e nas restantes não há qualquer erro. Obtenha destes dados um intervalo de confiança aproximado a 94% para $\lambda$ e interprete o significado do coeficiente 94%. (1.5)

Pelo T. L. C. temos que $\frac{\bar{X}-E[X]}{\sqrt{Var[X]/100}} = \frac{\bar{X}-\lambda}{\sqrt{\lambda/100}} \overset{a}{\sim} N(0,1)$. Tendo em conta que $E[X] = \lambda$, podemos substituir $\lambda$ no denominador da variável anterior por um seu estimador que seja, em particular, centrado como, por exemplo, $\bar{X}$.

Como $P\left(-c < \frac{\bar{X}-\lambda}{\sqrt{\bar{X}/100}} < c\right) \approx 0.94$ com $c = \phi^{-1}(0.97) = 1.8808$, tem-se $IAC_{\approx 0.94}(\lambda) = \left]\bar{X} - 1.8808\sqrt{\bar{X}/100}, \bar{X} + 1.8808\sqrt{\bar{X}/100}\right[$.

Na amostra observada tem-se $\bar{x} = \frac{25+16+6}{100} = 0.47$ obtendo-se $IC_{\approx 0.94}(\lambda) = ]0.341, 0.599[$.

O nível de confiança 94% pode ser interpretado como a percentagem de IC que se podem obter como concretizações do IAC que contêm o valor desconhecido de $\lambda$. Este coeficiente é assim uma medida da confiança que podemos ter que o IC calculado cumpra esse propósito.

| **Grupo II** | 5 valores |
|---|---|

**1.** Um material de revestimento usado em construção civil é vendido em cinco cores: verde, castanho, azul, preto e branco. Num estudo de mercado para analisar a popularidade das diferentes cores recolheu-se uma amostra casual referente a 300 vendas com os seguintes resultados:

| verde | castanho | azul | preto | branco |
|---|---|---|---|---|
| 55 | 35 | 88 | 54 | 68 |

Averigue se os dados colectados fornecem evidência suficiente para rejeitar a hipótese de as cinco cores serem igualmente preferidas pelos compradores. (2.0)

Numerando as cores de 1 a 5 pela ordem apresentada, seja $p_i$ a probabilidade de a cor $i$ ser escolhida por um comprador, com $i = 1, \ldots, 5$. Pretende-se testar $H_0 : \forall i : p_i = 1/5$ contra $H_1 : \exists i : p_i \neq 1/5$. Sob $H_0$ tem-se $e_i = 300p_i^0 = 60$, $\forall i$ e $Q^2 = \sum_{i=1}^{5} \frac{(O_i - E_i)^2}{E_i} \overset{a}{\sim} \chi^2_{(4)}$.
Valor-p=$P(Q^2 \geq 25.57) \approx 3.86 \times 10^{-5}$. Rejeita-se $H_0$ para os n. s. habituais, ou seja, há forte evidência para afirmar que a preferência dos compradores não se distribui uniformemente pelas 5 cores.

**2.** Numa certa análise económica estudou-se a relação entre o rendimento *per capita* mensal ($x$) e a despesa mensal em bens e serviços culturais ($Y$) de agregados familiares, ambos em dezenas de euros. Os dados relativos a 14 agregados familiares foram resumidos nas seguintes somas:

$$\sum_{i=1}^{14} x_i = 2910; \quad \sum_{i=1}^{14} x_i^2 = 985100; \quad \sum_{i=1}^{14} y_i = 379; \quad \sum_{i=1}^{14} y_i^2 = 15673.$$

Sob a validade do modelo de regressão linear $Y_i \sim N(\beta_0 + \beta_1 x_i, \sigma^2)$, $i = 1, \ldots, 14$, independentes, responda às questões seguintes, sabendo que a equação de regressão estimada é $\widehat{E}(Y \mid x) = 6.3412 + 0.0997\,x$.

(a) Interprete as estimativas dos parâmetros do modelo no contexto em análise e sugira um modelo de regressão que possa ser mais razoável para este caso. (1.5)

Estima-se que deve haver um aumento de 1 euro na despesa média em bens e serviços culturais pelo aumento de 10 euros no rendimento mensal e que um agregado familiar sem rendimentos gasta, em média, 63.412 euros em bens e serviços culturais. Esta segunda estimativa não parece ter sentido pelo que talvez seja preferível adoptar um MRLS sem ordenada na origem.

(b) Teste a hipótese de o rendimento *per capita* mensal não influenciar a despesa em bens e serviços culturais sob o modelo adoptado. (1.5)

Pretende-se testar $H_0 : \beta_1 = 0$ contra $H_1 : \beta_1 \neq 0$ com base em $T_0 = \frac{\hat{\beta}_1}{\sqrt{\frac{\hat{\sigma}^2}{\sum x_i^2 - n\bar{x}^2}}} \overset{H_0}{\sim} t_{(n-2)}$.

Tem-se $\hat{\sigma}^2 \approx 136.11$, $t_0 \approx 5.27$ e valor-p=$2P(T_0 \geq 5.27) = 2(1 - F_{t_{(12)}}(5.27)) \approx 2 \times 10^{-4}$. Rejeita-se $H_0$ aos n. s. habituais, pelo que não parece haver razão para duvidar de que o rendimento mensal influencia a despesa em bens e serviços culturais.