Probabilidades e Estatística

LEAN, LEE, LEGI, LEMat, LERC, LMAC, MEAer, MEAmbi, MEBiol, MEBiom, MEMec, MEQ

2º semestre – 2012/2013
14/06/2013 – 09:00

Duração: 90 minutos **2º teste A**

Justifique convenientemente todas as respostas!

Grupo I 10 valores

1. Uma concretização duma amostra aleatória de dimensão 100, da duração das chamadas feitas a partir de certo telemóvel (X, em minutos), permitiu concluir que a duração total das chamadas foi 320 minutos. Admitindo que X tem distribuição exponencial de parâmetro λ:

(a) Deduza o estimador de máxima verosimilhança de λ. (3.0)

$\mathscr{L}(\lambda; x_1,\ldots,x_n) = \prod_{i=1}^n f_X(x_i;\lambda) = \prod_{i=1}^n \lambda e^{-\lambda x_i} = \lambda^n e^{-\lambda \sum_{i=1}^n x_i}$
$\log \mathscr{L}(\lambda; x_1,\ldots,x_n) = n\log\lambda - \lambda\sum_{i=1}^n x_i$ (diferenciável em ordem a λ em $I\!R^+$)
$\frac{d\log\mathscr{L}(\lambda;x_1,\ldots,x_n)}{d\lambda} = 0 \iff \frac{n}{\lambda} - \sum_{i=1}^n x_i = 0 \iff \lambda = \bar{x}^{-1}$
$\frac{d^2\log\mathscr{L}(\lambda;x_1,\ldots,x_n)}{d\lambda^2} = -\frac{n}{\lambda^2} < 0,\ \forall\lambda\in I\!R^+$.
$\therefore \hat{\lambda}_{MV} = \bar{X}^{-1}$

(b) Obtenha, justificando, as estimativas de máxima verosimilhança da duração média de uma chamada e da probabilidade de uma chamada durar mais de 10 minutos. (2.0)

Pretende-se estimar $g(\lambda) = E[X] = \frac{1}{\lambda}$ e $h(\lambda) = P(X > 10) = \int_{10}^{+\infty} \lambda e^{-\lambda x}\,dx = e^{-10\lambda}$. Pela invariância dos estimadores de máxima verosimilhança tem-se as estimativas $\hat{g}_{MV}(\lambda) = g(\hat{\lambda}_{MV}) = \bar{x} = 3.2$ e $\hat{h}_{MV}(\lambda) = h(\hat{\lambda}_{MV}) = e^{-\frac{10}{\bar{x}}} = e^{-\frac{10}{3.2}} \approx 0.044$.

2. Para analisar a diferença (X, em milhares de euros) entre o volume de negócios dos meses de Janeiro de 2013 e Janeiro de 2012, em estabelecimentos de restauração de uma dada região, foram seleccionados ao acaso 61 restaurantes dessa região, tendo-se obtido os resultados seguintes: $\sum_{i=1}^{61} x_i = 135$ e $\sum_{i=1}^{61} x_i^2 = 325$. Supondo que X tem distribuição normal:

(a) Obtenha um intervalo de confiança a 98% para a variância de X. (2.5)

Sejam $Q = \frac{60S^2}{\sigma^2} \sim \chi^2_{(60)}$, $a = F^{-1}_{\chi^2_{(60)}}(0.01) = 37.48$ e $b = F^{-1}_{\chi^2_{(60)}}(0.99) = 88.38$ tais que $P(a \le Q \le b) = 0.98$.
$37.48 \le \frac{60S^2}{\sigma^2} \le 88.38 \iff \frac{60S^2}{88.38} \le \sigma^2 \le \frac{60S^2}{37.48} \implies IAC_{0.98}(\sigma^2) = \left[\frac{60S^2}{88.38}, \frac{60S^2}{37.48}\right]$
$s^2 = \frac{\sum_{i=1}^{61}(x_i-\bar{x})^2}{60} = \frac{\sum_{i=1}^{61} x_i^2 - 61\bar{x}^2}{60} = 0.4372 \implies IC_{0.98}(\sigma^2) = [0.297, 0.700]$

(b) Para o nível de significância de 1%, teste a hipótese de que, em média, não existiu diferença entre o volume de negócios nos dois referidos meses em estabelecimentos de restauração daquela região. (2.5)

$H_0: \mu = 0$ contra $H_1: \mu \neq 0$.
Seja $T = \frac{\bar{X}-\mu}{\sqrt{\frac{S^2}{61}}} \sim t_{(60)}$. Sob H_0 obtemos a estatística do teste, $T_0 = \frac{\bar{X}}{\sqrt{\frac{S^2}{61}}} \sim t_{(60)}$.
Para $\alpha = 0.01$ deve rejeitar-se H_0 se $|T_0| > F^{-1}_{t_{(60)}}(0.995) = 2.660$.
$t_0 = \frac{135/61}{\sqrt{\frac{0.4372}{61}}} = 26.143$.
Como t_0 pertence à região de rejeição então H_0 é rejeitada para $\alpha = 0.01$.
Alternativa: valor-p$= 2(1 - F_{t_{(60)}}(26.143)) \approx 0 < \alpha = 0.01$.

Grupo II	10 valores

1. Um grupo de 162 estudantes deslocados que frequentam o Ensino Superior foi seleccionado ao acaso e questionado sobre o número de horas mensal despendido a cozinhar, tendo-se registado os seguintes resultados:

Nº mensal de horas	$[0;5[$	$[5;10[$	$[10;15[$	$[15;+\infty[$
Nº de estudantes	73	57	16	16

(a) Teste a hipótese de que o número de horas mensal que os estudantes deslocados a frequentar o Ensino Superior despendem a cozinhar segue a distribuição indicada na tabela seguinte, para um nível de significância de 5%. (3.0)

Nº mensal de horas	$[0;5[$	$[5;10[$	$[10;15[$	$[15;+\infty[$
Probabilidade	0.40	0.35	0.15	0.10

Seja X ="número de horas mensal despendidas a cozinhar" e $p_i = P(X \in \text{Classe}_i)$, $i = 1,\ldots,4$. Pretende-se testar $H_0 : p_i = p_i^0,\ \forall i$ contra $H_1 : \exists i : p_i \neq p_i^0$

i	Classe_i	o_i	p_i^0	$e_i = np_i^0$
1	$[0;5[$	73	0.40	64.8
2	$[5;10[$	57	0.35	56.7
3	$[10;15[$	16	0.15	24.3
4	$[15;+\infty[$	16	0.10	16.2
		$n = 162$		

Como não é necessário agrupar classes ($k = 4$) nem há qualquer parâmetro estimado ($\beta = 0$), a estatística de teste é $Q_0 = \sum_{i=1}^{4} \frac{(O_i - E_i)^2}{E_i} \underset{H_0}{\overset{a}{\sim}} \chi^2_{(3)}$.

Para $\alpha = 0.05$ deve rejeitar-se H_0 se $Q_0 > F^{-1}_{\chi^2_{(3)}}(0.95) = 7.815$.

Como $q_0 = 3.877$ não pertence à região de rejeição então não se rejeita H_0 para $\alpha = 0.05$.

Alternativa: valor$-p = 1 - F_{\chi^2_{(3)}}(3.877) = 0.275 > 0.05$.

(b) Obtenha o valor-p do teste referido em (a) e comente (caso não consiga calcular exactamente o valor-p indique um intervalo onde este esteja contido). (1.0)

valor-p $= P(Q_0 > q_0 \mid H_0) = 1 - F_{\chi^2_{(3)}}(3.877) = 0.275$. H_0 deve ser rejeitada para níveis de significância superiores ou iguais a 0.275 e não deve ser rejeitada no caso contrário.

Nota: com as tabelas tem-se $1 - F_{\chi^2_{(3)}}(4.642) = 0.2 <$valor-p$< 0.3 = 1 - F_{\chi^2_{(3)}}(3.665)$.

2. Com o intuito de relacionar a carga movimentada Y (em centenas de toneladas) com o número x de contentores trabalhados por dia num porto, foram efectuadas observações durante 30 dias escolhidos ao acaso que conduziram a:

$$\sum_{i=1}^{30} x_i = 1220 \quad \sum_{i=1}^{30} x_i^2 = 49833 \quad \sum_{i=1}^{30} y_i = 3430 \quad \sum_{i=1}^{30} y_i^2 = 392613 \quad \sum_{i=1}^{30} x_i y_i = 139610$$

(a) Ajuste um modelo de regressão linear simples de Y sobre x e obtenha uma estimativa para o incremento esperado na carga movimentada resultante de um acréscimo de 5 contentores trabalhados num dia. (2.0)

$\hat{\beta}_1 = \frac{\sum_{i=1}^{n} x_i y_i - n\bar{x}\bar{y}}{\sum_{i=1}^{n} x_i^2 - n\bar{x}^2} = \frac{139610 - 1220 \times 3430/30}{49833 - 1220^2/30} = 0.5615$

$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} = 3430/30 - 0.5615 \times 1220/30 = 91.501$

$\hat{E}[Y|x] = 91.501 + 0.5615x$

Seja $\gamma = E[Y|x+5] - E[Y|x] = \beta_0 + \beta_1(x+5) - (\beta_0 + \beta_1 x) = 5\beta_1$. Então, $\hat{\gamma} = 5\hat{\beta}_1 = 2.807$.

(b) Calcule o coeficiente de determinação do modelo ajustado e comente o valor obtido. (1.0)

$$R^2 = \frac{\left(\sum_{i=1}^{n} x_i y_i - n\bar{x}\bar{y}\right)^2}{\left(\sum_{i=1}^{n} x_i^2 - n\bar{x}^2\right) \times \left(\sum_{i=1}^{n} y_i^2 - n\bar{y}^2\right)} = 0.154.$$

Apenas 15% da variabilidade da carga movimentada é explicada pelo número de contentores trabalhados. Este valor é muito baixo o que evidencia o mau ajustamento do modelo.

(c) Assumindo as hipóteses de trabalho habituais, teste a significância do modelo de regressão linear, usando um nível de significância de 1%. Relacione com o valor obtido na alínea anterior. (3.0)

Pretende-se testar $H_0 : \beta_1 = 0$ contra $H_1 : \beta_1 \neq 0$ com base em $T_0 = \frac{\hat{\beta}_1}{\sqrt{\frac{\hat{\sigma}^2}{\sum x_i^2 - 30\bar{x}^2}}} \overset{H_0}{\sim} t_{(28)}$.

Tem-se $\hat{\sigma}^2 \approx 13.118$, $t_0 \approx 2.298$ e valor-p=$2P(T_0 > 2.298) = 2(1 - F_{t_{(28)}}(2.298)) \approx 0.029 > 0.01$. Não se rejeita H_0 ao n. s. de 0.01, ou seja, sob o modelo adoptado, o número de contentores trabalhados por dia no porto não explica adequadamente a carga movimentada nesse dia. O mesmo é evidenciado pelo baixo valor de R^2 obtido na alínea anterior.