Probabilidades e Estatística

LEIC-A, LEIC-T, LMAC, LEMat, MA, MEAmbi, MEBiol, MEFT, LEGM, MEQ

2º semestre – 2014/2015
09/06/2015 – 11:00

Duração: 90 minutos **2º teste B**

Justifique convenientemente todas as respostas!

Grupo I 10 valores

1. Considere que a concretização de uma amostra aleatória de $X \sim Exp(\lambda)$ conduziu ao valor observado $\sum_{i=1}^{15} x_i = 45$.

(a) Deduza o estimador de máxima verosimilhança de λ. (2.5)

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

(b) Determine a estimativa de máxima verosimilhança da mediana de X. (1.5)

$F_X(x) = 0.5 \iff \int_0^x \lambda e^{-\lambda t}\,dt = 0.5 \iff x = \frac{\log 2}{\lambda}$.
Pretende-se estimar $g(\lambda) = \frac{\log 2}{\lambda}$. Pela invariância dos estimadores de máxima verosimilhança tem-se $\hat{g}_{MV}(\lambda) = g(\hat{\lambda}_{MV}) = \bar{X}\log 2$.
Com a amostra observada temos a estimativa $\hat{g}_{MV}(\lambda) = 3\log 2 \approx 2.079$.

(c) Tendo em conta que $2\lambda\sum_{i=1}^n X_i \sim \chi^2_{(2n)}$, obtenha o intervalo de confiança bilateral para o valor esperado de X com um nível de confiança de 0.9. (2.0)

Sejam $Q = 2\lambda\sum_{i=1}^{15} X_i \sim \chi^2_{(30)}$, $a = F^{-1}_{\chi^2_{(30)}}(0.05) = 18.49$ e $b = F^{-1}_{\chi^2_{(30)}}(0.95) = 43.77$.
Tem-se $P(a \le Q \le b) = 0.90 \iff P\left(\frac{2\sum_{i=1}^{15} X_i}{b} \le \frac{1}{\lambda} \le \frac{2\sum_{i=1}^{15} X_i}{a}\right) = 0.90$.
$\therefore IAC_{0.90}(1/\lambda) = \left[\frac{2\sum_{i=1}^{15} X_i}{43.77}, \frac{2\sum_{i=1}^{15} X_i}{18.49}\right]$.
Para a amostra observada $IC_{0.90}(1/\lambda) = [2.056, 4.867]$.

2. A duração (em horas) de certo tipo de bateria possui distribuição que se admite normal com valor esperado e variância desconhecidos. Um revendedor adquiriu recentemente um grande lote dessas baterias e registou os tempos de vida de 10 baterias, escolhidas ao acaso, tendo obtido $(x_1,\dots,x_{10})$ tal que $\sum_{i=1}^{10} x_i = 1989.78$ e $\sum_{i=1}^{10} x_i^2 = 485846.58$. O revendedor pretende verificar se o valor médio da duração não é superior a 200 horas. Teste esta hipótese e decida com base no valor-p do teste. (4.0)

$H_0: \mu \le 200$ contra $H_1: \mu > 200$.
Seja $T = \frac{\bar{X}-\mu}{\sqrt{\frac{S^2}{n}}} \sim t_{(n-1)}$. Sob H_0 obtemos a estatística do teste, $T_0 = \frac{\bar{X}-200}{\frac{S}{\sqrt{10}}} \overset{\mu=200}{\sim} t_{(9)}$.
Para a amostra observada tem-se $s = 99.96$ e $t_0 \approx -0.0323$.
Valor-p $= P(T_0 > -0.0323 \mid H_0) = 1 - F_{t_{(9)}}(-0.0323) \approx 0.5125$. Deve-se rejeitar H_0 para níveis de significância ≥ 0.5125 e não rejeitar no caso contrário. Para os níveis de significância usuais, $\alpha \in [0.01, 0.1]$, não há evidência suficiente para rejeitar H_0.

Grupo II	10 valores

1. Num estudo sobre a disponibilidade de medicamentos, registou-se o número de farmácias que foi necessário visitar até se encontrar um certo medicamento. (4.5)

Nº de farmácias visitadas	1	2	3	4
Frequência	26	12	10	12

Utilizando os dados acima tabelados, avalie ao nível de significância de 0.01 a hipótese de o número de farmácias visitadas até à compra do medicamento seguir uma distribuição geométrica de valor esperado 2.

Pretende-se testar $H_0 : X \sim Geo(1/2)$ contra $H_1 : X \not\sim Geo(1/2)$.
Começemos por juntar às 4 classes na tabela uma classe para valores de X superiores a 4 com frequência observada nula. Seja $p_i^0 = P(X \in \text{Classe}_i \mid H_0)$, $i = 1,\dots,5$.
Tem-se $p_i^0 = f_X(i)$, $i = 1,\dots,4$ e $p_5^0 = 1 - F_X(4)$.

Classe_i	o_i	p_i^0	$e_i = np_i^0$
1	26	0.5000	30.00
2	12	0.2500	15.00
3	10	0.1250	7.50
4	12	0.0625	3.75
≥ 5	0	0.0625	3.75
	$n = 60$		

Como as 2 últimas classes têm uma frequência esperada inferior a 5 é necessário agrupá-las ($k = 4$) e, como nenhum parâmetro foi estimado ($\beta = 0$), a estatística de teste é $Q_0 = \sum_{i=1}^{4} \frac{(O_i - E_i)^2}{E_i} \underset{H_0}{\overset{a}{\sim}} \chi^2_{(3)}$.
Tem-se $q_0 \approx 4.67$ e valor$-p = P(Q_0 > q_0 \mid H_0) = 1 - F_{\chi^2_{(3)}}(4.67) \approx 0.1976$. Como $\alpha = 0.01 <$valor$-p$, não há evidência suficiente para rejeitar H_0 ao n.s. de 0.01.

2. Uma amostra selecionada ao acaso de 100 alunos do 9º ano foi usada para comparar os resultados da classificação interna (x) com os obtidos no exame nacional (Y), tendo-se obtido os seguintes resultados:

$$\sum_{i=1}^{100} x_i = 920, \quad \sum_{i=1}^{100} y_i = 900, \quad \sum_{i=1}^{100} x_i^2 = 8\,600, \quad \sum_{i=1}^{100} y_i^2 = 8\,500, \quad \sum_{i=1}^{100} x_i y_i = 8\,400.$$

Considere válido o modelo de regressão linear simples, com as hipóteses de trabalho habituais.

(a) Estime a reta de regressão de Y sobre x e a variância de Y. (2.0)

$\hat{\beta}_1 = \frac{\sum_{i=1}^{n} x_i y_i - n\bar{x}\bar{y}}{\sum_{i=1}^{n} x_i^2 - n\bar{x}^2} = 0.8823$
$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} = 0.8828$
$\hat{E}[Y|x] = \hat{\beta}_0 + \hat{\beta}_1 x$, $x \in [0, 20]$.
$\hat{\sigma}^2 \approx 3.001$

(b) Calcule um intervalo de confiança a 90% para o declive da recta de regressão. O que pode concluir sobre a significância do modelo de regressão usado, ao nível de significância de 10%? (3.5)

Sejam $T = \frac{\hat{\beta}_1 - \beta_1}{\sqrt{\frac{\hat{\sigma}^2}{\sum x_i^2 - 100\bar{x}^2}}} \sim t_{(98)}$ e $a = F^{-1}_{t_{(98)}}(0.95) = 1.660$

$$P(-a \leq T \leq a) = 0.90 \iff P\left(\hat{\beta}_1 - a\sqrt{\frac{\hat{\sigma}^2}{\sum x_i^2 - 100\bar{x}^2}} \leq \beta_1 \leq \hat{\beta}_1 + a\sqrt{\frac{\hat{\sigma}^2}{\sum x_i^2 - 100\bar{x}^2}}\right) = 0.90$$

$\text{IAC}_{0.90}(\beta_1) = \left[\hat{\beta}_1 - 1.660\sqrt{\frac{\hat{\sigma}^2}{\sum x_i^2 - 100\bar{x}^2}}, \hat{\beta}_1 + 1.660\sqrt{\frac{\hat{\sigma}^2}{\sum x_i^2 - 100\bar{x}^2}}\right]$
$\text{IC}_{0.90}(\beta_1) = [0.6357, 1.1289]$
A um nível de significância de 0.1 pode-se concluir que há uma relação de tipo linear entre x e Y ($\beta_1 \neq 0$) uma vez que $0 \notin IC_{0.90}(\beta_1)$