Probabilidades e Estatística

LEGM, LEIC-A, LEIC-T, LMAC, LETI, MEC, MEFT

1º semestre – 2014/2015
08/01/2015 – 11:00

Duração: 90 minutos **2º teste B**

Justifique convenientemente todas as respostas!

Grupo I 10 valores

1. Seja $(X_1, X_2, \ldots, X_n)$ uma amostra aleatória de uma população com distribuição dependente de um parâmetro desconhecido, θ. Considere T_1 e T_2, estimadores de θ, com valores esperados e variâncias dados por: $E(T_1) = \theta + \frac{1}{n}$, $Var(T_1) = \frac{1}{n^2}$, $E(T_2) = \theta$ e $Var(T_2) = \frac{3}{n^2}$. Determine qual dos estimadores mencionados é mais eficiente para estimar o parâmetro θ. (2.0)

$\frac{EQM[T_1]}{EQM[T_2]} = \frac{\text{Var}[T_1] + (\text{E}[T_1] - \theta)^2}{\text{Var}[T_2] + (\text{E}[T_2] - \theta)^2} = \frac{2/n^2}{3/n^2} = \frac{2}{3} < 1 \implies T_1$ é um estimador de θ mais eficiente que T_2.

2. Considere a variável aleatória $X \sim Poi(\lambda)$, a qual modela o número de participações de sinistros autómoveis a determinada seguradora num período de uma hora, e uma amostra aleatória $(X_1, X_2, \ldots, X_n)$ de X.

(a) Determine o estimador de máxima verosimilhança do valor esperado do número de participações de sinistros automóveis à seguradora numa hora. (2.5)

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

(b) Calcule a estimativa de máxima verosimilhança da probabilidade de ocorrerem mais de 3 participações de sinistros automóveis à seguradora numa hora, sabendo que a concretização $(x_1, x_2, \ldots, x_{20})$ de uma amostra aleatória de dimensão 20 de X conduziu a $\sum_{i=1}^{20} x_i = 40$. (1.0)

Pretende-se estimar $g(p) = P(X > 3) = 1 - F_{Poi(\lambda)}(3)$. Pela invariância dos estimadores de invariância tem-se que $\hat{g}_{MV}(p) = g\left(\hat{p}_{MV}\right) = 1 - F_{Poi(\bar{X})}(3)$.
Com $\bar{x} = 2$ tem-se $\hat{P}(X > 3) = 1 - F_{Poi(2)}(3) = 0.1429$.

3. O tempo (em minutos) de instalação de um programa estatístico em certo tipo de computadores possui distribuição normal com desvio padrão 0.3 minutos. O fabricante do programa afirmou que o valor esperado do tempo de instalação mencionado é 2 minutos. Com a finalidade de testar a veracidade da afirmação do fabricante, um técnico informático efetuou 10 instalações independentes do programa, em computadores do tipo referido, tendo obtido um tempo médio de instalação de 2.07 minutos.

(a) Teste, ao nível de significância de 5%, a validade da afirmação feita pelo fabricante. (3.0)

$H_0 : \mu = 2$ contra $H_1 : \mu \neq 2$.
Seja $T = \frac{\bar{X} - \mu}{\sqrt{\frac{\sigma^2}{n}}} \sim N(0,1)$. Sob H_0 obtemos a estatística do teste, $T_0 = \frac{\bar{X} - 2}{\frac{0.3}{\sqrt{10}}} \sim N(0,1)$.
Para $\alpha = 0.05$ deve rejeitar-se H_0 se $|T_0| > \Phi^{-1}(0.975) = 1.96$.
$t_0 = \frac{0.07}{\frac{0.3}{\sqrt{10}}} \approx 0.7379$.
Como t_0 não pertence à região de rejeição então H_0 não é rejeitada para $\alpha = 0.05$.
Alternativa: valor-p$= 2(1 - \Phi(0.7379)) \approx 0.4606 > \alpha = 0.05$.

(b) Calcule a probabilidade de o teste estatístico usado na alínea anterior, aplicado a uma nova amostra de dimensão 10, indicar que o fabricante não tem razão caso o valor esperado do tempo de instalação do programa nos computadores em questão seja efetivamente 2.5 minutos. (1.5)

Pretende-se $P(|T_0| > 1.96 \mid \mu = 2.5)$. Como $\mu = 2.5$ tem-se agora que $T^* = \frac{\bar{X}-2.5}{\frac{0.3}{\sqrt{10}}} \sim N(0,1)$.
$P(|T_0| > 1.96 \mid \mu = 2.5) = 1 - P(|T_0| \leq 1.96 \mid \mu = 2.5) = 1 - P(-7.27 \leq T^* \leq -3.27) = 1 - (\Phi(-3.27) - \Phi(-7.27)) \approx 0.9995$.

Grupo II 10 valores

1. A observação, durante um determinado período de tempo, das entradas de clientes numa grande superfície comercial com 4 portas de acesso conduziu aos seguintes resultados:

Porta	Norte	Sul	Este	Oeste
Nº Entradas	327	402	351	380

Teste a hipótese de as entradas de clientes na superfície comercial mencionada se distribuírem uniformemente pelas quatro portas de acesso. Decida com base no valor-p. (4.0)

Numerando as portas de 1 a 4, seja X =“porta escolhida por um cliente”. Pretende-se testar $H_0 : X \sim U(\{1,\dots,4\})$ contra $H_1 : X \not\sim U(\{1,\dots,4\})$.
Seja $p_i^0 = P(X = i \mid H_0) = 0.25$, $i = 1,\dots,4$.

i	o_i	p_i^0	$e_i = np_i^0$
1	327	0.25	365
2	402	0.25	365
3	351	0.25	365
4	380	0.25	365
	$n = 1460$		

Como todas as classes têm uma frequência esperada superior a 5 não é necessário agrupar classes ($k = 4$) e, não havendo qualquer parâmetro estimado ($\beta = 0$), a estatística de teste é $Q_0 = \sum_{i=1}^{4} \frac{(O_i-E_i)^2}{E_i} \underset{H_0}{\overset{a}{\sim}} \chi^2_{(3)}$.
Tem-se $q_0 \approx 8.86$ e valor$-p = P(Q_0 > q_0 \mid H_0) = 1 - F_{\chi^2_{(3)}}(8.86) = 0.0312$. Deve-se rejeitar H_0 para níveis de significância ≥ 0.0312 e não rejeitar no caso contrário.

2. Considere que o modelo de regressão linear, $Y = \beta_0 + \beta_1 x + \varepsilon$, com as suposições de trabalho habituais, é adequado para explicar a relação entre a variável aleatória Y e a variável x. Com base numa amostra $\{(x_i, y_i)\}_{i=1,2,\cdots,12}$, com $x_i \in [1,12]$, obteve-se a seguinte estimativa da reta de regressão de mínimos quadrados, com arredondamentos a quatro casas decimais: $\widehat{E(Y|x)} = 0.9523 - 0.9788\,x$. Sabe-se também que:

$$\sum_{i=1}^{12} x_i = 78, \quad \sum_{i=1}^{12} x_i^2 = 650, \quad \sum_{i=1}^{12} y_i = -64.92, \quad \sum_{i=1}^{12} y_i^2 = 488.3406, \quad \sum_{i=1}^{12} x_i y_i = -561.95.$$

(a) Determine o coeficiente de determinação do modelo e comente o resultado obtido. (2.0)

$$R^2 = \frac{\left(\sum_{i=1}^{n} x_i y_i - n\bar{x}\bar{y}\right)^2}{\left(\sum_{i=1}^{n} x_i^2 - n\bar{x}^2\right) \times \left(\sum_{i=1}^{n} y_i^2 - n\bar{y}^2\right)} \approx 0.99913.$$

Conclui-se que 99.9% da variabilidade observada em Y é explicada pelo MRLS o que evidencia o bom ajustamento desse modelo aos dados.

(b) Obtenha um intervalo de confiança a 90% para o valor esperado de Y quando $x = 6$. (4.0)

$E[Y|x=6]=\beta_0+6\beta_1$

Sejam $T=\frac{(\hat\beta_0+6\hat\beta_1)-(\beta_0+6\beta_1)}{\sqrt{\left(\frac{1}{12}+\frac{(\bar x-6)^2}{\sum x_i^2-12\bar x^2}\right)\hat\sigma^2}}\sim t_{(10)}$ e $a=F^{-1}_{t_{(10)}}(0.95)=1.812$

$P(-a\le T\le a)=0.9\iff$

$P\left(\hat\beta_0+6\hat\beta_1-a\sqrt{\left(\frac{1}{12}+\frac{(\bar x-6)^2}{\sum x_i^2-12\bar x^2}\right)\hat\sigma^2}\le\beta_0+6\beta_1\le\hat\beta_0+6\hat\beta_1+a\sqrt{\left(\frac{1}{12}+\frac{(\bar x-6)^2}{\sum x_i^2-12\bar x^2}\right)\hat\sigma^2}\right)=0.9$

$\mathrm{IAC}_{0.9}(\beta_0+6\beta_1)=\left[\hat\beta_0+6\hat\beta_1-1.812\sqrt{\left(\frac{1}{12}+\frac{(\bar x-6)^2}{\sum x_i^2-12\bar x^2}\right)\hat\sigma^2},\hat\beta_0+6\hat\beta_1+1.812\sqrt{\left(\frac{1}{12}+\frac{(\bar x-6)^2}{\sum x_i^2-12\bar x^2}\right)\hat\sigma^2}\right]$

$\hat\beta_0+6\hat\beta_1=-4.9205$

$\sqrt{\left(\frac{1}{12}+\frac{(\bar x-6)^2}{\sum x_i^2-12\bar x^2}\right)\hat\sigma^2}=0.03226$

$\mathrm{IC}_{0.9}(\beta_0+6\beta_1)=[-4.979,-4.8620]$