Probabilidades e Estatística

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

1º semestre – 2013/2014
09/01/2014 – 09:00

Duração: 90 minutos **2º teste A**

Justifique convenientemente todas as respostas!

Grupo I 10 valores

1. Seja $(X_1, X_2, \ldots, X_n)$ uma amostra aleatória proveniente de uma população $X \sim \text{Normal}(0,\ \theta)$, onde $\theta = \sigma^2 = Var(X) > 0$.

(a) Deduza o estimador de máxima verosimilhança do parâmetro θ. (3.5)

$\mathcal{L}(\theta; x_1, \ldots, x_n) = \prod_{i=1}^n f_X(x_i;\theta) = \prod_{i=1}^n (2\pi\theta)^{-1/2} e^{-x_i^2/2\theta} = (2\pi\theta)^{-n/2} e^{-\sum_{i=1}^n x_i^2/2\theta}$

$\log \mathcal{L}(\theta; x_1, \ldots, x_n) = -\frac{n}{2}\log(2\pi\theta) - \sum_{i=1}^n x_i^2/2\theta$ (diferenciável em ordem a θ em $I\!R^+$)

$\frac{d\log\mathcal{L}(\theta;x_1,\ldots,x_n)}{d\lambda} = 0 \iff -\frac{n}{2\theta} + \sum_{i=1}^n \frac{x_i^2}{2\theta^2} = 0 \iff \theta = \sum_{i=1}^n \frac{x_i^2}{n}$

$\left.\frac{d^2\log\mathcal{L}(\theta;x_1,\ldots,x_n)}{d\lambda^2}\right|_{\theta=\sum_{i=1}^n \frac{x_i^2}{n}} = \left.\frac{n}{2\theta^2} - \sum_{i=1}^n \frac{x_i^2}{\theta^3}\right|_{\theta=\sum_{i=1}^n \frac{x_i^2}{n}} = -\frac{n^3}{2\left(\sum_{i=1}^n x_i^2\right)^2} < 0.$

$\therefore \hat{\theta}_{MV} = \sum_{i=1}^n \frac{X_i^2}{n}$

(b) Averigue se $T = \frac{1}{n}\sum_{i=1}^n X_i^2$ é um estimador centrado para o parâmetro θ. (1.5)

$E[T] = E\left[\frac{1}{n}\sum_{i=1}^n X_i^2\right] = \frac{\sum_{i=1}^n E[X_i^2]}{n} = E\left[X^2\right] = Var[X] + (E[X])^2 = \theta \implies T$ é um estimador centrado de θ.

2. Um fabricante de máquinas de calcular conjectura que no máximo 1% da sua produção é defeituosa. Com base numa amostragem aleatória, foram observadas 1000 calculadoras tendo-se verificado que 12 eram defeituosas.

(a) Teste, ao nível de significância de 5%, a hipótese colocada pelo fabricante. (3.0)

A amostra observada é uma concretização de uma amostra aleatória de $X \sim Ber(p)$, em que $p = P(\text{peça defeituosa})$. Quer-se testar $H_0: p \le 0.01$ contra $H_1: p > 0.01$.
Uma vez que o tamanho da amostra é suficientemente grande temos, pelo TLC, $Z = \frac{\bar{X}-E[X]}{\sqrt{\frac{Var[X]}{n}}} = \frac{\bar{X}-p}{\sqrt{\frac{p(1-p)}{n}}} \overset{a}{\sim} N(0,1)$. Sob H_0, admitindo $p = 0.01$, obtemos a estatística do teste, $Z_0 = \frac{\bar{X}-0.01}{\sqrt{\frac{0.01\times 0.99}{n}}} \overset{a}{\sim} N(0,1)$.
Para $\alpha = 0.05$ deve rejeitar-se H_0 se $Z_0 > \Phi^{-1}(0.95) = 1.6449$. Para a amostra observada temos $\bar{x} = 12/1000 = 0.012$ e $z_0 = 0.6356$. Como z_0 não pertence à região de rejeição então H_0 não é rejeitada para $\alpha = 0.05$.
Alternativa: valor$-p = 1-\Phi(0.6356) = 0.2625 > 0.05$.

(b) No caso de haver de facto 2% de calculadoras defeituosas, que dimensão mínima deve ter a amostra para que o teste efectuado na alínea anterior tenha uma probabilidade aproximada de erro de tipo II (isto é, de incorrectamente não rejeitar H_0) que não exceda 5%? (2.0)

Quer-se determinar o valor mínimo de n tal que $P\left(Z_0 \le 1.6449 \,\middle|\, p = 0.02\right) \le 0.05$. Dado que $p = 0.02$, temos agora que $Z^* = \frac{\bar{X}-0.02}{\sqrt{\frac{0.02\times 0.98}{n}}} \overset{a}{\sim} N(0,1)$.

$P\left(Z_0 \le 1.6449 \,\middle|\, p = 0.02\right) = P\left(Z^* \le \frac{1.6449\sqrt{0.01\times 0.99} - 0.01\sqrt{n}}{\sqrt{0.02\times 0.98}}\right) \approx \Phi\left(\frac{1.6449\sqrt{0.01\times 0.99} - 0.01\sqrt{n}}{\sqrt{0.02\times 0.98}}\right) \le 0.05 \iff$
$\frac{1.6449\sqrt{0.01\times 0.99} - 0.01\sqrt{n}}{\sqrt{0.02\times 0.98}} \le \Phi^{-1}(0.05) = -1.6449 \iff n \ge 1551.98.$
$\therefore n = 1552.$

Grupo II	10 valores

1. A tabela seguinte indica, num conjunto de 200 dias escolhidos ao acaso, o número de pessoas que às 18h30m aguardam atendimento numa certa farmácia: (4.0)

Nº de pessoas	[0,2]	[3,7]	[8,10]
Nº de dias	30	140	30

Teste a hipótese de a variável aleatória X, que indica o número de pessoas que num dia de funcionamento da farmácia aguardam atendimento às 18h30m, seguir uma distribuição de Poisson de valor esperado igual a 5 pessoas. Calcule, justificando, o valor-p do teste e decida com base no valor obtido, tendo em conta os níveis de significância usuais.

Seja X ="número de pessoas que num dia de funcionamento da farmácia aguardam atendimento às 18h30m". Pretende-se testar $H_0 : X \sim Poi(5)$ contra $H_1 : X \nsim Poi(5)$.

$$\text{Seja } p_i^0 = P(X \in \text{Classe}_i \mid H_0) = \begin{cases} P(X \in [0,2] \mid H_0) = F_{Poi(5)}(2) = 0.1247, & i=1 \\ P(X \in [3,7] \mid H_0) = F_{Poi(5)}(7) - F_{Poi(5)}(2) = 0.7419, & i=2 \\ P(X \in [8,10] \mid H_0) = F_{Poi(5)}(10) - F_{Poi(7)}(7) = 0.1197, & i=3 \\ P(X \in [11,+\infty] \mid H_0) = 1 - F_{Poi(5)}(10) = 0.0137, & i=4 \end{cases}.$$

i	Classe_i	o_i	p_i^0	$e_i = np_i^0$
1	$[0,2]$	30	0.1247	24.94
2	$[3,7]$	140	0.7419	148.38
3	$[8,10]$	30	0.1197	23.94
4	$[11,+\infty[$	0	0.0137	2.74
		$n=200$		

Como 25% das classes têm uma frequência esperada inferior a 5, é necessário agrupar as duas últimas classes ($k=3$) e, não havendo qualquer parâmetro estimado ($\beta=0$), a estatística de teste é $Q_0 = \sum_{i=1}^{3} \frac{(O_i-E_i)^2}{E_i} \underset{H_0}{\overset{a}{\sim}} \chi^2_{(2)}$.
Tem-se $q_0 = 1.9130$ e valor$-p = P(Q_0 > q_0 \mid H_0) = 1 - F_{\chi^2_{(2)}}(1.9130) = 0.3842$. Deve-se rejeitar H_0 para níveis de significância ≥ 0.3842 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.

2. Para estudar o efeito da viscosidade de um tipo óleo (x, em Ns/m^2) no desgaste de uma peça (Y, em 10^{-4} mm^3) selecionou-se uma amostra ao acaso de 7 peças onde foi aplicado esse tipo de óleo. As observações conduziram aos seguintes resultados:

$$\sum_{i=1}^{7} x_i = 147 \qquad \sum_{i=1}^{7} y_i = 1164 \qquad \sum_{i=1}^{7} x_i^2 = 4324.42 \qquad \sum_{i=1}^{7} y_i^2 = 206088 \qquad \sum_{i=1}^{7} x_i y_i = 20724.9$$

Admita que o modelo de regressão linear simples é adequado para explicar a relação existente entre x e Y.

(a) Indique esse modelo, especifique os pressupostos necessários para que ele tenha validade estatística e obtenha as estimativas de mínimos quadrados dos coeficientes da recta de regressão. (2.0)

Admita-se que $Y_i \sim N(\beta_0 + \beta_1 x_i, \sigma^2)$, $i = 1, \ldots, 7$, com Y_i e Y_j não correlacionadas $\forall i \neq j$.
$\hat{\beta}_1 = \frac{\sum_{i=1}^{n} x_i y_i - n\bar{x}\bar{y}}{\sum_{i=1}^{n} x_i^2 - n\bar{x}^2} = -2.99$
$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} = 229.10$.

(b) Obtenha o valor ajustado para o desgaste esperado numa peça quando se aplica óleo com viscosidade igual a 22 Ns/m^2. Calcule, ainda, o resíduo associado à seguinte observação $(x, y) = (22, 172)$. (1.0)

$\hat{E}[Y|x=22] = \hat{\beta}_0 + 22\hat{\beta}_1 = 163.29$. Resíduo $= y - \hat{E}[Y|x=22] = 8.71$

(c) Deduza um intervalo de confiança a 90% para a ordenada na origem da recta de regressão. (3.0)

Sejam $T = \frac{\hat{\beta}_0 - \beta_0}{\sqrt{\left(\frac{1}{7} + \frac{\bar{x}^2}{\sum x_i^2 - 7\bar{x}^2}\right)\hat{\sigma}^2}} \sim t_{(5)}$ e $a = F^{-1}_{t_{(5)}}(0.95) = 2.015$

$P(-a \le T \le a) = 0.90 \Leftrightarrow P\left(\hat{\beta}_0 - a\sqrt{\left(\frac{1}{7} + \frac{\bar{x}^2}{\sum x_i^2 - 7\bar{x}^2}\right)\hat{\sigma}^2} \le \beta_0 \le \hat{\beta}_0 + a\sqrt{\left(\frac{1}{7} + \frac{\bar{x}^2}{\sum x_i^2 - 7\bar{x}^2}\right)\hat{\sigma}^2}\right) = 0.90$

$\text{IAC}_{0.90}(\beta_0) = \left[\hat{\beta}_0 - a\sqrt{\left(\frac{1}{7} + \frac{\bar{x}^2}{\sum x_i^2 - 7\bar{x}^2}\right)\hat{\sigma}^2}, \hat{\beta}_0 + a\sqrt{\left(\frac{1}{7} + \frac{\bar{x}^2}{\sum x_i^2 - 7\bar{x}^2}\right)\hat{\sigma}^2}\right]$

$\hat{\sigma}^2 = \frac{1}{n-2}\left[\left(\sum_{i=1}^{n} y_i^2 - n\bar{y}^2\right) - \left(\hat{\beta}_1\right)^2\left(\sum_{i=1}^{n} x_i^2 - n\bar{x}^2\right)\right] = 292.30.$

$\therefore \text{IC}_{0.90}(\beta_0) = [204.75, 253.44]$