DM DEPARTAMENTO DE MATEMÁTICA TÉCNICO LISBOA

Probabilidades e Estatística

LEGI, LEGM, LMAC, MEAmbi, MEAer, MEC

1º semestre – 2016/2017
12/01/2017 – **11:00**

Duração: 90 minutos **2º teste**

Justifique convenientemente todas as respostas!

Grupo I	10 valores

1. Num estudo relativo ao tempo de resposta de determinado sistema operativo concluiu-se que o tempo de resposta, em segundos, é modelado por uma variável aleatória X com distribuição exponencial de parâmetro λ, onde $\lambda > 0$ é desconhecido.

(a) Admita que a concretização de uma amostra aleatória de dimensão 20 da variável aleatória X conduziu a $\sum_{i=1}^{20} x_i = 80$. Determine as estimativas de máxima verosimilhança do parâmetro λ e da probabilidade do tempo de resposta do sistema ser superior a 5 segundos. (3.5)

$\mathscr{L}(\lambda; x_1, \ldots, x_{20}) \overset{iid}{=} \prod_{i=1}^{20} f_X(x_i;\lambda) = \prod_{i=1}^{20} \lambda e^{-\lambda x_i} = \lambda^{20} e^{-\lambda \sum_{i=1}^{20} x_i}$
$\log \mathscr{L}(\lambda; x_1, \ldots, x_{20}) = 20 \log\lambda - \lambda \sum_{i=1}^{20} x_i$ (diferenciável em ordem a λ em $I\!R^+$)
$\frac{d \log \mathscr{L}(\lambda; x_1,\ldots,x_{20})}{d\lambda} = 0 \iff \frac{20}{\lambda} - \sum_{i=1}^{20} x_i = 0 \iff \lambda = \bar{x}^{-1}$
$\frac{d^2 \log \mathscr{L}(\lambda; x_1,\ldots,x_{20})}{d\lambda^2} = -\frac{20}{\lambda^2} < 0, \ \forall \lambda \in I\!R^+$.
$\therefore \hat{\lambda}_{MV} = \bar{X}^{-1}$.
Pretende-se também estimar $g(\lambda) = P(X > 5) = \int_5^{+\infty} \lambda e^{-\lambda t}\, dt = e^{-5\lambda}$. Pela invariância dos estimadores de máxima verosimilhança tem-se $\hat{g}_{MV}(\lambda) = g(\hat{\lambda}_{MV}) = e^{-5/\bar{X}}$.
Com a amostra observada temos as estimativas $\hat{\lambda}_{MV} = 1/4$ e $\hat{g}_{MV}(\lambda) = e^{-5/4} \approx 0.2865$.

(b) $\bar{X}$ é um estimador centrado de $1/\lambda$? Justifique. (1.5)

Sim porque $E[\bar{X}] = E\left[\frac{\sum_{i=1}^{20} X_i}{20}\right] = \frac{\sum_{i=1}^{20} E[X_i]}{20} = E[X] = \frac{1}{\lambda}$.

2. Admite-se que o número de acidentes de trabalho por ano numa certa indústria é descrito pela variável aleatória X. Os registos referentes aos últimos 40 anos para essa indústria conduziram a uma média amostral de 2.5 acidentes por ano e um desvio padrão amostral de 1.5 acidentes por ano.

(a) Teste a hipótese do valor esperado de X ser igual a 2, a um nível de significância aproximado de 5%. (3.0)

$H_0 : \mu = 2$ contra $H_1 : \mu \neq 2$.
Seja $T = \frac{\bar{X}-\mu}{\sqrt{\frac{S^2}{n}}} \overset{a}{\sim} N(0,1)$. Sob H_0 obtemos a estatística do teste, $T_0 = \frac{\bar{X}-2}{\sqrt{\frac{S^2}{n}}} \underset{H_0}{\overset{a}{\sim}} N(0,1)$.
Para $\alpha = 0.05$ deve rejeitar-se H_0 se $|T_0| > \Phi^{-1}(0.975) = 1.96$.
Como $t_0 \approx 2.108$ pertence à região de rejeição então H_0 é rejeitada para $\alpha = 0.05$.

(b) Assumindo que X tem distribuição de Poisson de parâmetro λ, $\lambda > 0$, determine um intervalo de confiança aproximado a 95% para λ. Considere a variável aleatória fulcral $Z = (\bar{X} - \lambda)/\sqrt{\bar{X}/n}$, cuja distribuição é aproximadamente Normal(0,1). (2.0)

Sendo $\Phi^{-1}(0.975) = 1.96$, tem-se
$P(-1.96 \leq Z \leq 1.96) \approx 0.95 \iff P\left(\bar{X} - 1.96\sqrt{\bar{X}/n} \leq \lambda \leq \bar{X} + 1.96\sqrt{\bar{X}/n}\right) \approx 0.95$.
$\therefore IAC_{\approx 0.95}(\lambda) = \left[\bar{X} - 1.96\sqrt{\bar{X}/n}, \bar{X} + 1.96\sqrt{\bar{X}/n}\right]$.
Para a amostra observada $IC_{\approx 0.95}(\lambda) = [2.010, 2.990]$.

Grupo II	10 valores

1. Um fabricante de chocolate vende o seu produto em quatro embalagens distintas: A, B, C, D. Depois de inquirir 600 potenciais consumidores do seu chocolate em relação à embalagem que prefeririam comprar, obteve os dados no quadro seguinte

Embalagem	A	B	C	D
Nº vendas	146	151	152	151

Teste a hipótese de as vendas se distribuirem uniformente pelos quatro tipos de embalagens em que o chocolate é comercializado. Decida com base no valor-p. (4.0)

Numerando os tipos de embalagem de 1 a 4, seja X ="embalagem preferida por um consumidor". Pretende-se testar $H_0 : X \sim U(\{1,\ldots,4\})$ contra $H_1 : X \not\sim U(\{1,\ldots,4\})$.
Seja $p_i^0 = P(X = i \mid H_0) = 1/4$, $i = 1,\ldots,4$.

i	o_i	p_i^0	$e_i = np_i^0$
1	146	0.25	150
2	151	0.25	150
3	152	0.25	150
4	151	0.25	150
	$n = 600$		

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 $q_0 = 11/75 \approx 0.1467$ e valor$-p = P(Q_0 > q_0 \mid H_0) = 1 - F_{\chi^2_{(3)}}(0.1467) \approx 0.9974$. Deve-se rejeitar H_0 para níveis de significância ≥ 0.9974 e não rejeitar no caso contrário. Para os níveis de significância usuais os dados não fornecem evidência para a rejeição de H_0.

2. O modelo de regressão linear simples foi usado para estudar a relação entre o tempo, x, decorrido depois da inoculação de uma vacina em cinco indivíduos diferentes e a contagem bacteriana, Y, neles observada. Obtiveram-se os seguintes resultados:

Dias após inoculação da vacina, x	3	6	7	8	9
Contagem bacteriana, y (em milhares)	92	131	201	231	330

$$\bar{x} = 6.6, \quad \bar{y} = 197, \quad \sum_{i=1}^5 x_i^2 = 239, \quad \sum_{i=1}^5 y_i^2 = 228287, \quad \sum_{i=1}^5 x_i y_i = 7287.$$

(a) Obtenha as estimativas dos mínimos quadrados dos coeficientes da reta de regressão. (2.0)

$\hat{\beta}_1 = \frac{\sum_{i=1}^n x_i y_i - n\bar{x}\bar{y}}{\sum_{i=1}^n x_i^2 - n\bar{x}^2} = 36.075$ e $\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} = -47.698$

(b) Após ter enunciado as hipóteses de trabalho que entender mais convenientes, construa um intervalo de confiança a 90% para o declive da recta de regressão. (3.0)

Admite-se que as variáveis $Y_i = Y \mid x = x_i$ são não correlacionadas e que $Y_i \sim N\left(\beta_0 + \beta_1 x_i, \sigma^2\right)$ para $i = 1,\ldots,5$.
Sejam $T = \frac{\hat{\beta}_1 - \beta_1}{\sqrt{\frac{\hat{\sigma}^2}{\sum x_i^2 - 5\bar{x}^2}}} \sim t_{(3)}$ e $a = F^{-1}_{t_{(3)}}(0.95) = 2.353$.

$$P(-a \leq T \leq a) = 0.90 \iff P\left(\hat{\beta}_1 - a\sqrt{\frac{\hat{\sigma}^2}{\sum x_i^2 - 5\bar{x}^2}} \leq \beta_1 \leq \hat{\beta}_1 + a\sqrt{\frac{\hat{\sigma}^2}{\sum x_i^2 - 5\bar{x}^2}}\right) = 0.90$$

$$\text{IAC}_{0.90}(\beta_1) = \left[\hat{\beta}_1 - 2.353\sqrt{\frac{\hat{\sigma}^2}{\sum x_i^2 - 5\bar{x}^2}}, \hat{\beta}_1 + 2.353\sqrt{\frac{\hat{\sigma}^2}{\sum x_i^2 - 5\bar{x}^2}}\right]$$

Calculando $\hat{\sigma}^2 = 1700.226$ obtem-se $IC_{0.90}(\beta_1) = [16.000, 58.151]$

(c) Estime o valor esperado da contagem bacteriana num indivíduo, após 7 dias da inoculação da vacina. (1.0)

Pretende-se estimar $E[Y|x=7]=\beta_0+7\beta_1$.
$\hat{E}[Y|x=7]=\hat{\beta}_0+7\hat{\beta}_1=211.830$