Probabilidades e Estatística

LEIC-A, LEIC-T, LEGM, MEEC, MEFT

2º semestre – 2012/2013
14/06/2013 – 11:00

Duração: 90 minutos **2º teste B**

Justifique convenientemente todas as respostas!

Grupo I 10 valores

1. Considere uma população X com distribuição normal de valor médio nulo e variância θ e uma amostra aleatória de X, de dimensão n, $X_1,\dots,X_n$.

(a) Obtenha o estimador de máxima verosimilhança de θ e calcule a respectiva estimativa para a amostra: -0.71, 0.22, 0.04, 1.15. (3.0)

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

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

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

$\frac{d^2\log\mathcal{L}(\theta;x_1,\dots,x_n)}{d\theta^2} = \frac{n}{2\theta^2} - \frac{\sum_{i=1}^n x_i^2}{\theta^3}$

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

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

Para a amostra observada tem-se $\hat\theta_{MV} = \frac{1.8766}{4} = 0.4692$.

(b) Decida se o estimador $T = \frac{1}{n}\sum_{i=1}^n X_i^2$ é ou não centrado para θ. (2.0)

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

2. Uma estação agronómica desenvolveu um novo tipo de macieiras que foram vendidas tanto para o continente europeu como para o americano. A estação agronómica decidiu analisar a produtividade em dois países destes continentes. Para tal seleccionou ao acaso 100 árvores em França e 80 no Canadá, e contabilizou a sua produção unitária (produção por árvore), a qual se assume seguir uma distribuição normal. Os valores da produção unitária das amostras observadas conduziram aos resultados seguintes:

- França: média de 50 Kg, com um desvio padrão de 10 Kg.
- Canadá: média de 54 Kg, com um desvio padrão de 12.5 Kg.

(a) Construa um intervalo de confiança a 95% para o valor esperado da produção unitária deste novo tipo de macieiras quando plantado em França. (2.0)

Sejam $X_1 (X_2)$ ="produção unitária em França (no Canadá)", $T = \frac{\bar X_1 - \mu_1}{\sqrt{\frac{S_1^2}{100}}} \sim t_{(99)}$ e $a = F^{-1}_{t_{(99)}}(0.975) =$ 1.984 tal que $P(-a \le T \le a) = 0.95$.

$-1.984 \le \frac{\bar X_1 - \mu_1}{\sqrt{\frac{S_1^2}{100}}} \le 1.984 \iff \bar X_1 - 1.984\frac{S_1}{10} \le \mu_1 \le \bar X_1 + 1.984\frac{S_1}{10} \implies IAC_{0.95}(\mu_1) = \left[\bar X_1 - 1.984\frac{S_1}{10}, \bar X_1 + 1.984\frac{S_1}{10}\right]$

$\therefore IC_{0.95}(\mu_1) = [48.016, 51.984]$

(b) Decida, com base no valor-p, se a produtividade média é idêntica em ambos os países. (3.0)

$H_0: \mu_1 - \mu_2 = 0$ contra $H_1: \mu_1 - \mu_2 \neq 0$.
Seja $T = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} \overset{a}{\sim} N(0,1)$. Sob H_0 obtemos a estatística do teste, $T_0 = \frac{(\bar{X}_1 - \bar{X}_2)}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} \overset{a}{\sim} N(0,1)$.
Para um dado α deve rejeitar-se H_0 se $|T_0| > \Phi^{-1}(1 - \alpha/2)$ (teste bilateral).
$t_0 = \frac{50-54}{\sqrt{\frac{10^2}{100} + \frac{12.5^2}{80}}} = -2.328$.
valor-p$= 2\Phi(-2.328) \approx 0.02 \implies H_0$ deve ser rejeitada para n. s. superiores ou iguais a 0.02 e não deve ser rejeitada no caso contrário.

Grupo II — 10 valores

1. A análise do número de golos por jogo, em 30 jogos escolhidos ao acaso de um campeonato de futebol, conduziu aos seguintes resultados:

Nº de golos	0	1	2	3	4 ou mais
Nº de jogos	1	7	11	5	6

(a) Admitindo que o número de golos por jogo desse campeonato tem distribuição de Poisson (de parâmetro λ), e que a estimativa de máxima verosimilhança do parâmetro λ é 2.4, estime a probabilidade de haver mais do que 3 golos num jogo. (1.0)

Sejam $X =$"número de golos por jogo", com $X \sim Pois(\lambda)$, e $g(\lambda) = P(X > 3) = 1 - P(X \leq 3) = 1 - F_{Pois(\lambda)}(3)$.
Pela invariância dos estimadores de máxima verosimilhança temos $\hat{g}_{MV}(\lambda) = g(\hat{\lambda}_{MV}) = 1 - F_{Pois(2.4)}(3) = 1 - 0.7787 = 0.2213$.

(b) Teste ao nível de significância de 10% a hipótese de o número de golos por jogo nesse campeonato ter distribuição de Poisson de parâmetro 2.4. (3.0)

$H_0: X \sim Pois(2.4)$ contra $H_1: X \nsim Pois(2.4)$
Estatística de teste: $Q_0 = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i} \underset{H_0}{\overset{a}{\sim}} \chi^2_{(k-\beta-1)}$
Sejam $p_i^0 = P(X = i|H_0) = e^{-2.4}\frac{2.4^i}{i!}$, $i = 0,1,2,3$ e $p_4^0 = P(X \geq 4|H_0) = 1 - \sum_{i=0}^{3} p_i^0 = 0.2213$

	o_i	p_i^0	$e_i = np_i^0$
0	1	0.0907	2.721
1	7	0.2177	6.531
2	11	0.2613	7.839
3	5	0.2090	6.270
≥ 4	6	0.2213	6.639
	$n = 30$		

Não é necessário agrupar classes ($k = 5$) nem há qualquer parâmetro estimado ($\beta = 0$).
Para $\alpha = 0.10$ deve rejeitar-se H_0 se $Q_0 > F^{-1}_{\chi^2_{(4)}}(0.90) = 7.779$
Como $q_0 = 2.716$ não pertence à região de rejeição então não se rejeita H_0 para $\alpha = 0.10$.
Alternativa: valor$-p = 1 - F_{\chi^2_{(4)}}(2.716) = 0.6065 > 0.10$.

2. Dados relativos ao peso (x) de 16 animais domésticos e ao peso de ração (Y) por eles consumida diariamente, ambos em unidades convenientes, conduziram aos seguintes resultados:

$$\sum_{i=1}^{16} x_i = 144.5 \quad \sum_{i=1}^{16} x_i^2 = 2465.9 \quad \sum_{i=1}^{16} y_i = 23.8 \quad \sum_{i=1}^{16} y_i^2 = 58.1 \quad \sum_{i=1}^{16} x_i y_i = 375.8$$

(a) Ajuste um modelo de regressão linear simples de Y sobre x e obtenha uma estimativa para a diferença nos valores esperados de ração consumida entre dois animais cujos pesos difiram de 5 unidades. (2.0)

$\hat{\beta}_1 = \frac{\sum_{i=1}^n x_i y_i - n\bar{x}\bar{y}}{\sum_{i=1}^n x_i^2 - n\bar{x}^2} = \frac{375.8 - 144.5 \times 23.8}{2465.9 - 144.5^2/16} = 0.1386$

$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} = 23.8/16 - 0.1386 \times 144.5/16 = 0.2361$

$\hat{E}[Y|x] = 0.2361 + 0.1386x$

Seja $\gamma = E[Y|x+5] - E[Y|x] = \beta_0 + \beta_1(x+5) - (\beta_0 + \beta_1 x) = 5\beta_1$. Então, $\hat{\gamma} = 5\hat{\beta}_1 = 0.693$.

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

$$R^2 = \frac{\left(\sum_{i=1}^n x_i y_i - n\bar{x}\bar{y}\right)^2}{\left(\sum_{i=1}^n x_i^2 - n\bar{x}^2\right) \times \left(\sum_{i=1}^n y_i^2 - n\bar{y}^2\right)} = 0.9820.$$

Conclui-se que 98.2% da variabilidade da ração consumida diariamente é explicada pelo peso dos animais. Este valor é muito elevado o que evidencia o bom 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 $T_0 = \frac{\hat{\beta}_1}{\sqrt{\frac{\hat{\sigma}^2}{\sum x_i^2 - 16\bar{x}^2}}} \overset{H_0}{\sim} t_{(14)}$.

Tem-se $\hat{\sigma}^2 \approx 0.0272$, $t_0 \approx 28.602$ e valor-p=$2P(T_0 > 28.602) = 2(1 - F_{t_{(14)}}(28.602)) \approx 8 \times 10^{-14} < 0.01$. Rejeita-se H_0 ao n. s. de 0.01, ou seja, sob o modelo adoptado, o peso dos animais explica adequadamente o peso de ração consumida por dia. O mesmo é evidenciado pelo elevado valor de R^2 obtido na alínea anterior.