INSTITUTO SUPERIOR DE ECONOMIA E GESTÃO

ESTATÍSTICA II – LICENCIATURA EM GESTÃO

Época Normal – 7 de Janeiro de 2013

Parte Prática

Nome: ____ Nº________

Espaço reservado para classificações				
1. (20)	3. (15)	5a. (20)	5c. (20)	**T:**
2. (15)	4. (20)	5b. (10)	5d. (20)	**P:** _____

Em todos os testes de hipóteses que fizer, formule as hipóteses em teste, indique a estatística de teste e a sua distribuição. Para os intervalos de confiança proceda de forma semelhante para a variável fulcral. Se necessitar de espaço dispõe de uma página em branco no fim do enunciado

1. De uma população com distribuição exponencial foi observada uma amostra de dimensão 5 que apresentou os seguintes valores: 3, 5, 2, 9, 1. Atendendo a que função densidade de probabilidade da distribuição exponencial é $f(x \mid \lambda) = \lambda e^{-\lambda x}, x \geq 0$, obtenha uma estimativa para o parâmetro λ através do método da máxima verosimilhança. Apenas considere a condição de primeira ordem.

$$L(\lambda) = \prod_{i=1}^{n} \lambda e^{-\lambda x_i} = \lambda^n e^{-\lambda \sum_{i=1}^{n} x_i}$$

$$l(\lambda) = \log L(\lambda) = n \log \lambda - \lambda \sum_{i=1}^{n} x_i$$

$$\frac{dl(\lambda)}{d\lambda} = \frac{n}{\lambda} - \sum_{i=1}^{n} x_i$$

$$\frac{dl(\lambda)}{d\lambda} = 0 \iff \hat{\lambda} = \frac{n}{\sum_{i=1}^{n} x_i} = \frac{1}{\bar{x}}$$

Com base na amostra obtem-se a seguinte estimativa para λ: 5 / (3 + 5 + 2 + 9 + 1) = 0.25

2. Inquiriram-se 200 alunos, seleccionados aleatoriamente entre a população estudantil do ISEG, com o objectivo de saber se estavam inclinados em votar em determinada lista concorrente às eleições para a AEISEG. Desses, 120 responderam "sim". Com base nesta informação, construa um intervalo de confiança a 95% para a proporção de alunos do ISEG que estarão inclinados a votar nessa lista. Diga se o intervalo obtido garante a essa lista, com um grau de confiança de 95%, uma vitória por maioria absoluta.

Variável fulcral para proporção de Bernoulli:

$$\frac{\bar{X}-\theta}{\sqrt{\frac{\bar{X}(1-\bar{X})}{n}}} \overset{a}{\sim} N(0,1)$$

Intervalo de confiança para proporção de Bernoulli:

$$\left(\bar{x}-z_{\alpha/2}\sqrt{\frac{\bar{x}(1-\bar{x})}{n}}, \bar{x}+z_{\alpha/2}\sqrt{\frac{\bar{x}(1-\bar{x})}{n}}\right)$$

$\bar{x}$ = 120 / 200 = 0.6, $z_{0.025}$ = 1.96

Intervalo de confiança:

(0.6 – 1.96 . (0.6 (1 – 0.6) / 200)$^{1/2}$, 0.6 + 1.96 . ((1 – 0.6) / 200)$^{1/2}$) = (0.5321, 0.6679)

O intervalo de confiança não inclui o valor 0.5. A proporção observada na amostra sugere, com um grau de confiança de 95%, uma vitória por maioria absoluta.

3. O Director Comercial de uma grande empresa pretende comparar a volatilidade do volume de vendas mensal da filial do Algarve, *X*, com a da filial de Lisboa, *Y*. Ele suspeita que, devido a factores sazonais, a variância do volume de vendas mensal da filial do Algarve, σ_X^2, é *superior* à variância do volume de vendas mensal da filial de Lisboa, σ_Y^2. Para testar esta hipótese, o Director Comercial recolheu aleatoriamente os volumes de vendas em 16 meses, para cada uma das filiais. As variâncias corrigidas amostrais do volume de vendas das duas filiais foram calculadas, tendo obtido-se os seguintes valores: $s'^2_X = 42.5$ e $s'^2_Y = 16.8$. Com uma dimensão de teste de 0.05, pode afirmar-se que a suspeita do Director Comercial é correcta?

Pretende testar-se: $H_0: \sigma_X^2 = \sigma_Y^2$ contra $H_1: \sigma_X^2 > \sigma_Y^2$.

Estatística de teste: $F = \frac{S'^2_X}{S'^2_Y} \sim F(m-1, n-1)$ → f_{obs} = 42.5 / 16.8 = 2.53

F(15, 15, 0.05) = 2.40 → Região de rejeição: $W_F = \{f : f > 2.40\}$

O valor observado da estatística de teste encontra-se na região de rejeição. Assim, conclui-se pela rejeição da hipótese nula e aceitação da alternativa. O teste sugere que a suspeita do Director Comercial é correcta.

4. Um produtor de espectáculos musicais pretende seleccionar uma banda portuguesa para fazer a primeira parte de um concerto da banda britânica *Radiohead*. Ele pondera convidar a banda de Coimbra *Wraygunn* para fazer a primeira parte do concerto. No entanto, o produtor quer certificar-se de que os fans dos *Radiohead* têm tendência para também apreciar os *Wraygunn*. Para este fim, inquiriu aleatoriamente 200 pessoas. Entre as pessoas inquiridas, 92 afirmaram gostar das duas bandas, 53 afirmaram gostar apenas dos *Radiohead*, 11 afirmaram gostar apenas dos *Wraygunn*, e 44 afirmaram não gostar de nenhuma das duas bandas. Com uma dimensão de 0.05, teste se existe associação entre ser-se apreciador da banda *Radiohead* e ser-se apreciador da banda *Wraygunn*.

H_0 : para todo (i, j) $p_{ij} = p_{io}\, p_{oj}$ contra H_1 : existir (i, j) tal que $p_{ij} \neq p_{io}\, p_{oj}$.

Estatística de teste:

$$Q = \sum_{i=1}^{r}\sum_{j=1}^{s} \frac{\left(N_{ij} - \frac{N_{io}N_{oj}}{n}\right)^2}{\frac{N_{io}N_{oj}}{n}} \overset{a}{\sim} \chi^2\{(r-1)(s-1)\}$$

Frequências observadas

	sim	não	N_{io}
sim	**92**	**53**	145
não	**11**	**44**	55
N_{oj}	103	97	200

Frequências esperadas

	sim	não
sim	74.68	70.33
não	28.33	26.68

Desvios $(N_{ij} - N_{io}N_{oj} / n)^2 / (N_{io}N_{oj} / n)$

	sim	não
sim	4.02	4.27
não	10.60	11.25

$q_{obs} = 30.14$ graus de liberdade = $(2-1)(2-1) = 1$
$\alpha = 0.05$ $Q_{0.05} = 3.841$ Região de rejeição: $W_Q = \{ q : q > 3.841\}$
valor-p = $P(\chi^{(1)} > q_{obs} \mid H_0) = 4.03E-08$

Rejeita-se a hipótese nula de não haver associação entre ser-se apreciador dos *Radiohead* e ser-se apreciador dos *Wraygunn*.

5. Pretende-se efectuar um estudo sobre os níveis de desenvolvimento humano nos países do Mundo. Para tal, especificou-se o seguinte modelo:

$$idh_t = \beta_1 + \beta_2 esp\_med\_vida_t + \beta_3 escol\_med_t + \beta_4 pnb\_pc_t + u_t$$

Onde,

- idh – valor do índice de desenvolvimento humano;
- esp_med_vida – esperança média de vida à nascença, em anos;
- $escol\_med$ – escolaridade média, em anos;
- pnb_pc – produto nacional bruto *per capita*, em <u>milhares</u> de unidades monetárias.

Analisando um grupo composto por 187 países, estimou-se o modelo acima com o EXCEL, cujo resultado se encontra no **Modelo 1**, em anexo.

a) Interprete a estimativa para o coeficiente β_2 e analise a sua significância estatística com uma dimensão de teste de 0.05. Não se esqueça de apresentar o teste de hipóteses e a estatística de teste em questão.

$b_2 = 0.00790$ → se a esperança média de vida à nascença aumentar 1 ano, estima-se que o índice de desenvolvimento humano aumente, em média, 0.0079 valores, ceteris paribus.

$$H_0: \beta_2 = 0 \quad vs \quad H_1: \beta_2 \neq 0,$$

ET: $t_2 = \frac{b_2 - \beta_2^0}{s_{b_2}} \sim t(n-k) = t(163)$

Pelo output,

$t_{2,obs} = 19.58 \Rightarrow p_{obs} = 2P\left(t_2 \geq |t_{2,obs}| \middle| H_0\right) = 0.000 < \alpha = 0.05$

Logo, rejeita-se H_0 ao nível de 5%, o que nos leva a concluir que o regressor "esp_med_vida" é estatisticamente significativo.

b) Interprete o valor do coeficiente de determinação, R^2. Utilize-o para encontrar o valor do coeficiente de determinação ajustado, $\bar{R}^2$, em falta no quadro.

$R^2 = 0.96166$ → Este modelo explica cerca de 96% da variação total do índice de desenvolvimento humano.

$$\bar{R}^2 = 1 - (1 - R^2)\frac{n-1}{n-k} = 1 - (1 - 0.96166)\frac{187-1}{187-4} \approx 0.96103$$

c) Com o auxílio do **Modelo 2**, em anexo, teste a validade da seguinte restrição: $\beta_2 = 3\beta_4$. Admita uma dimensão de 0.05.

Pretende-se efectuar o seguinte teste de hipóteses:

$$H_0: \beta_2 - 3\beta_4 = 0 \text{ vs } H_1: \beta_2 - 3\beta_4 \neq 0$$

Estatística-teste: $F = \frac{(VR_0 - VR_1)/m}{VR_1/(n-k)} \sim F(m, n-k) = F(1{,}183)$

$f_{obs} = \dfrac{(0.21484 - 0.21087)/1}{0.21087/(187-4)} \approx 3.4453$

$W_F = \{f_{obs}: f_{obs} > 3.84\}$

$p_{obs} = P(F \geq f_{obs}|H_0) = P(F \geq 3.4453) \approx 0.06504 > \alpha = 0.05$

Logo, não se rejeita H_0 ao nível de 5% (a restrição é validada pelos dados).

d) Com base no **Modelo 1**, apresente um intervalo de previsão a 90% para o IDH de um país com as seguintes características: esperança média de vida à nascença de 79.5 anos, escolaridade média de 7.7 anos e produto nacional bruto *per capita* de 20573 unidades monetárias. Para tal, admita que $s\sqrt{1 + c(X^TX)^{-1}c^T} = 0.0342$.

<u>Previsão pontual</u>

IP a 90% para

$$\hat{y}_o = E(idh_t | esp\_med\_vida_t = 79.5, escol\_med_t = 7.7, pnb\_pc_t = 20.573)$$

VF: $t_d = \frac{\widehat{y_o} - \hat{y}_o}{s\sqrt{1 + c(X^TX)^{-1}c^T}} \sim t(n-k) = t(183)$

IP a 90% para y_0: $\left(\hat{y}_o \mp t_{0.05} s\sqrt{1 + c(X^TX)^{-1}c^T}\right)$

- $t_{0.05} \approx 1.645$
- $\hat{y}_o = -0.12548 + 0.0079 \times 79.5 + 0.02777 \times 7.7 + 0.00212 \times 20.573 \approx 0.76$
- $s\sqrt{1 + c(X^TX)^{-1}c^T} = 0.0342$

Logo,

IP a 90% para y_o: $(0.76 \mp 1.645 \times 0.0342) = (0.7037, 0.8163)$

Continuação da questão ______

Modelo 1

$$idh_t = \beta_1 + \beta_2 esp\_med\_vida_t + \beta_3 escol\_med_t + \beta_4 pnb\_pc_t + u_t$$

Regression Statistics	
Multiple R	0.98064
R Square	0.96166
Adjusted R Square	[redacted]
Standard Error	0.03395
Observations	187

ANOVA

	Df	*SS*	*MS*	*F*	*Significance F*
Regression	3	5.28887	1.76296	1529.95793	0.00000
Residual	183	0.21087	0.00115		
Total	186	5.49974			

	Coefficients	*Standard Error*	*t Stat*	*P-value*
Intercept	-0.12548	0.02250	-5.57722	0.00000
esp_med_vida	0.00790	0.00040	19.58426	0.00000
escol_med	0.02777	0.00125	22.19169	0.00000
pnb_pc	0.00212	0.00020	10.48946	0.00000

Modelo 2

$$idh_t = \alpha_1 + \alpha_2 escol\_med_t + \alpha_3 X_t + v_t,$$

Regression Statistics	
Multiple R	0.98027
R Square	0.96094
Adjusted R Square	0.96051
Standard Error	0.03417
Observations	187

ANOVA

	Df	*SS*	*MS*	*F*	*Significance F*
Regression	2	5.28490	2.64245	2263.16876	0.00000
Residual	184	0.21484	0.00117		
Total	186	5.49974			

	Coefficients	*Standard Error*	*t Stat*	*P-value*
Intercept	-0.09449	0.01517	-6.22701	0.00000
escol_med	0.02815	0.00124	22.65186	0.00000
X	0.00245	0.00009	26.57954	0.00000

NOTA: $X_t = 3 \times esp\_med\_vida_t + pnb\_pc_t$