Estatística II – Licenciatura em Gestão
EN - 17 Junho 2014

Parte teórica

Nome: ____ **Nº**________

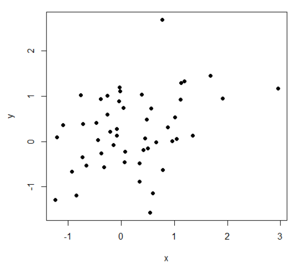
1. **Perguntas de Verdadeiro/Falso (1.5 valores) -** Para cada afirmação assinale se esta é verdadeira (V) ou falsa (F). Uma resposta certa vale 0.3 e uma resposta errada penaliza em idêntico valor.

	V	F
Num teste de dimensão $\alpha = 0.05$ em que o valor-*p*=0.07, rejeita-se H_0.		X
No MRL $y_t = \beta_1 + \beta_2 x_t + u_t$, $t = 1,2,\ldots,50$, em que o coeficiente de correlação amostral entre y_t e x_t é 0.75, pode-se garantir que, utilizando os mínimos quadrados, se obterá $b_2 > 0$	X	
Um intervalo de confiança a 95% para θ contém, com probabilidade 0.95, o valor de θ		X
Com base numa amostra casual de dimensão $n = 3$, o estimador $T = 0.5X_1 + 0.4X_2 + 0.1X_3$, sendo centrado é também o estimador mais eficiente para μ		X
Num teste de hipóteses H_0 e H_1 têm de constituir uma partição do espaço do parâmetro	X	

2. **Escolha Múltipla (2.25 valores) -** Para cada pergunta assinale com X a alternativa correcta. Uma resposta certa vale 0.75 valores e uma resposta errada penaliza em 0.25 valores.
   a. Quando se afirma que o modelo de regressão linear $y_t = \beta_1 + \beta_2 x_{t2} + \beta_3 x_{t3} + u_t$ com $t = 1,2,\ldots,n$ verifica a hipótese de homocedasticidade condicionada, isto significa que, para $t, s = 1,2,\ldots,n$ com $t \neq s$:
      - ☐ $\mathrm{cov}(u_t, u_s \mid X) \neq 0$
      - X $\mathrm{var}(u_t \mid X) = \sigma^2$
      - ☐ $\mathrm{cov}(u_t, u_s \mid X) = 0$
      - ☐ $\mathrm{var}(u_t \mid X) = \sigma^2 x_{t2}$

   b. Se um teste de hipótese simples contra hipótese simples tem potência 0.3 isto significa que
      - X Quando H_1 é verdadeira, a probabilidade de H_0 ser rejeitada é 0.3
      - ☐ Quando H_0 é verdadeira, a probabilidade de H_0 ser rejeitada é 0.3
      - ☐ Quando H_0 é verdadeira, a probabilidade de H_0 não ser rejeitada é 0.3
      - ☐ Quando H_1 é verdadeira, a probabilidade de H_0 não ser rejeitada é 0.3

   c. A figura representa o diagrama de dispersão dos valores y_t e x_t. Qual poderá ser o valor do coeficiente de determinação que se obtém ao estimar a regressão $y_t = \beta_1 + \beta_2 x_t + u_t$?



- ☐ -0.20
- ☐ -0.95
- X 0.20
- ☐ 0.99

3. Perguntas de desenvolvimento (2.25 valores) – alínea a) 1 valor; alínea b) 1.25 valores.

a. Sabendo que X tem como função de distribuição $F(x\,|\,\theta)=1-\left(\dfrac{\theta}{x+\theta}\right)^2$, $x>0,\ \theta>0$, um investigador deseja testar a hipótese $P(X>2)=0.64$. Tratar-se-á de uma hipótese paramétrica? Caso responda afirmativamente explicite a hipótese em causa, caso contrário explique porquê.

Sim, trata-se de uma hipótese paramétrica já que

$$P(X>2)=0.64 \Leftrightarrow 1-\left(1-\left(\frac{\theta}{2+\theta}\right)^2\right)=0.64 \Leftrightarrow \left(\frac{\theta}{2+\theta}\right)^2=0.64 \Leftrightarrow \frac{\theta}{2+\theta}=0.8 \Leftrightarrow \theta=8$$

Logo testar $H_0: P(X>2)=0.64$ é equivalente a testar $H_0:\theta=8$ que é indiscutivelmente uma hipótese paramétrica (a distribuição de X é conhecida)

b. Considere o MRL $y_t=\beta_1+\beta_2\,x_{t2}+\beta_3\,x_{t3}+\beta_4\,x_{t4}+u_t$ que foi estimado pelos MQ com base numa amostra de dimensão n. Obtenha a regressão auxiliar para testar $H_0:\begin{cases}\beta_2=2\,\beta_3\\ \beta_3+\beta_4=0\end{cases}$.

A regressão auxiliar é obtida introduzindo H_0 no modelo.

Como $H_0:\begin{cases}\beta_2=2\,\beta_3\\ \beta_3+\beta_4=0\end{cases} \Leftrightarrow H_0:\begin{cases}\beta_2=2\,\beta_3\\ \beta_4=-\beta_3\end{cases}$ vem

$$\begin{aligned}
y_t &= \beta_1+\beta_2\,x_{t2}+\beta_3\,x_{t3}+\beta_4\,x_{t4}+u_t\\
&= \beta_1+2\beta_3\,x_{t2}+\beta_3\,x_{t3}-\beta_3\,x_{t4}+u_t\\
&= \beta_1+\beta_3\left(2\,x_{t2}+x_{t3}-x_{t4}\right)+u_t\\
&= \beta_1+\beta_3\,x_t^*+u_t
\end{aligned}$$

com $x_t^*=2\,x_{t2}+x_{t3}-x_{t4}$. Estimada a regressão auxiliar pelos MQ procede-se a um teste F comparando esta regressão com a regressão original.

Estatística II – Licenciatura em Gestão
EN - 17 Junho 2014

Parte Prática

Nome: ____ **Nº**________

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

Em todos os testes de hipóteses que fizer, formule as hipóteses em teste, indique a estatística de teste e a sua distribuição. Se nada for dito em contrário utilize $\alpha = 0.05$. Para os intervalos de confiança proceda de forma semelhante para a variável fulcral.
Se necessitar de espaço dispõe de uma folha em branco no fim do enunciado

1) Assuma que o montante levantado numa máquina Mutibanco pode ser bem modelado por uma variável aleatória com função densidade dada por:

$$f(x|\theta) = \frac{x}{\theta}\exp\left\{-\frac{x^2}{2\theta}\right\}, \qquad (x > 0)$$

Com base na amostra aleatória $(X_1, X_2, \ldots, X_n)$, determine o estimador de máxima verosimilhança para θ.

$$L(\theta) = \prod_{i=1}^{n} \frac{x_i}{\theta}\exp\left(-\frac{x_i^2}{2\theta}\right) = \frac{\prod_{i=1}^{n} x_i}{\theta^n}\exp\left(-\frac{1}{2\theta}\sum_{i=1}^{n} x_i^2\right)$$

$$l(\theta) = \ln L(\theta) = \ln\left(\prod_{i=1}^{n} x_i\right) - n\ln\theta - \frac{1}{2\theta}\sum_{i=1}^{n} x_i^2$$

$$\frac{d\,\mathrm{l}(\theta)}{d\theta} = 0 \Leftrightarrow -\frac{n}{\theta} + \frac{2}{4\theta^2}\sum_{i=1}^{n} x_i^2 = 0 \Leftrightarrow \frac{\sum_{i=1}^{n} x_i^2}{2\theta^2} = \frac{n}{\theta} \Leftrightarrow \theta = \frac{\sum_{i=1}^{n} x_i^2}{2n}$$

$$\frac{d^2\,\mathrm{l}(\theta)}{d\theta^2} = \frac{n}{\theta^2} - \frac{4\theta}{4\theta^4}\sum_{i=1}^{n} x_i^2 = \frac{n}{\theta^2} - \frac{1}{\theta^3}2n\theta = -\frac{n}{\theta^2} < 0$$

Estimador MV para θ: $\hat{\theta} = \frac{\sum_{i=1}^{n} X_i^2}{2n}$

2) Com o intuito de estudar a abstenção nas próximas eleições legislativas recolheu-se uma amostra aleatória de 500 eleitores, dos quais 225 declararam que se iriam abster.

a. Construa um intervalo de confiança a 95% para a proporção de eleitores, na população, que tem intenção de se abster nas próximas eleições legislativas.

$VF: Z = \frac{\bar{X}-\theta}{\sqrt{\frac{\bar{X}(1-\bar{X})}{n}}} \dot{\sim} N(0,1)$ $\quad 1-\alpha = 0.05$ logo $z_{\alpha/2} = 1.96$ $\quad \bar{x} = 225/500 = 0.45$

IC a 95% para θ: $\left(\bar{x} - 1.96\sqrt{\frac{\bar{x}(1-\bar{x})}{n}}, \bar{x} + 1.96\sqrt{\frac{\bar{x}(1-\bar{x})}{n}}\right)$=

$$= (0.45 - 0.044, 0.45 + 0.044) = (0.406, 0.494)$$

b. Um comentador político afirma que a percentagem de abstenção será no máximo 40%. Com base num teste de hipóteses comente a afirmação deste comentador.

$$H_0: \theta \leq 0.4, \qquad H_1: \theta > 0.4$$

$$\text{Como } Z = \frac{\bar{X}-\theta}{\sqrt{\frac{\theta(1-\theta)}{n}}} \dot{\sim} N(0,1) \text{ tem-se, sob } H_0,\ ET: Z = \frac{\bar{X}-0.4}{\sqrt{\frac{0.4(1-0.4)}{n}}} \dot{\sim} N(0,1)$$

$$Z_{obs} = 2.282$$

$$p_{obs} = P(Z \geq 2.282) = 1 - \Phi(2.282) = 0.011$$

Dado o valor-p bastante abaixo do nível de significância pretendido ($\alpha = 0.05$), rejeitamos H_0, isto é rejeitamos a afirmação do comentador político.

3) O dono de um ginásio pretende saber se as modalidades praticadas pelos clientes estão relacionadas com as suas idades. Para isso observou as preferências de alguns dos seus clientes:

Idade	Modalidade		
	Musculação	Pilates	Dança
35 ou menos	150	150	100
Mais de 35	150	250	200

É de concluir que as preferências se podem considerar independentes da idade?

$$H_0: \forall(i,j): p_{ij} = p_{io}p_{oj}, \qquad (i = 1,2 \wedge j = 1,2,3)$$

$$H_1: \exists(i,j): p_{ij} \neq p_{io}p_{oj}, \qquad (i = 1,2 \wedge j = 1,2,3)$$

$$ET: Q = \sum_{i=1}^{r}\sum_{j=1}^{s}\frac{\left(N_{ij} - fe_{ij}\right)^2}{fe_{ij}} \dot{\sim} \chi_{\{(r-1)(s-1)\}}$$

Calculo das frequências esperadas:

$$fe_{11} = \frac{n_{1o}n_{o1}}{n} = \frac{400 \times 300}{1000} = 120$$

$$fe_{12} = \frac{n_{1o}n_{o2}}{n} = \frac{400 \times 400}{1000} = 160$$

$$fe_{13} = \frac{n_{1o}n_{o3}}{n} = \frac{400 \times 300}{1000} = 120$$

$$fe_{21} = \frac{n_{2o}n_{o1}}{n} = \frac{600 \times 300}{1000} = 180$$

$$fe_{22} = \frac{n_{2o}n_{o2}}{n} = \frac{600 \times 400}{1000} = 240$$

$$fe_{23} = \frac{n_{2o}n_{o3}}{n} = \frac{600 \times 300}{1000} = 180$$

$$Q_{obs} = \frac{(150-120)^2}{120} + \frac{(150-160)^2}{160} + \frac{(100-120)^2}{120} + \frac{(150-180)^2}{180} + \frac{(250-240)^2}{240} + \frac{(200-180)^2}{180}$$

$$= 7.5 + 0.625 + 3.333 + 5 + 0.417 + 2.222 = 19{,}097$$

$$P_{obs} = P(Q \geq 19{,}097) \approx 0$$

Assim, rejeita-se H_0 a qualquer nível de significância habitual. Logo, com os dados de que dispomos, podemos concluir que as preferências dos utentes daquele ginásio não são independentes da sua idade.

4) Com o intuito de estudar o salário dos gestores empresariais, foi recolhida uma amostra de 177 gestores, tendo-se estimado o seguinte modelo:

$$lsal_t = \beta_1 + \beta_2 lucros_t + \beta_3 lvendas_t + \beta_4 exper_t + \beta_5 exper2_t + u_t$$

Onde:

- sal – Salário do gestor, em milhares de euros;
- $lsal$ – Logaritmo do salário;
- $lucros$ – Lucros da empresa em milhões de euros;
- $vendas$ – Vendas da empresa em milhões de euros;
- $lvendas$ – Logaritmo das vendas;
- $exper$ – Número de anos como gestor da empresa;
- $exper2$ – Quadrado de $exper$.

No **Anexo 1** figuram os resultados obtidos na estimação deste modelo.

a. Interprete as estimativas obtidas para os coeficientes β_2 e β_3 e teste a significância estatística de cada um destes parâmetros.

$b_2 = 0.0002$: Estima-se que, em média, o aumento de 1 milhão de euros nos lucros da empresa, *ceteris paribus*, levará a um aumento de, aproximadamente, 0.02% no salário do gestor.

$b_3 = 0.196$: Estima-se que, em média, o aumento de 1% das vendas da empresa, *ceteris paribus*, levará a um aumento de, aproximadamente, 0.196% no salário do gestor.

$$H_0: \beta_j = 0, \qquad H_1: \beta_j \neq 0, \qquad j = 2{,}3$$

$$ET: T_{b_j} = \frac{b_j - \beta_j^0}{se_{b_j}} \sim t_{(n-k)} \qquad \text{com } n=177 \text{ e } k=5$$

$$T_{b_2}^{obs} \approx 1.566, \qquad T_{b_3}^{obs} \approx 5.925$$

$$P_{b_2}^{obs} \approx 0.119, \qquad P_{b_3}^{obs} \approx 0.000$$

Tendo em conta os níveis habituais de significância, concluímos que a variável *lucros* não é estatisticamente significativa, ao contrário das *vendas.*

b. Interprete o valor do coeficiente de determinação, e utilize-o para encontrar o valor do coeficiente de determinação ajustado, $\bar{R}^2$, em falta no **Anexo 1**.

$R^2 = 0.334 \rightarrow$ Este modelo explica cerca de 33.4% da variação total do logaritmo do salário do gestor.

$$\bar{R}^2 = 1 - (1 - R^2)\frac{n-1}{n-k}, \qquad \bar{R}^2 = 1 - (1 - 0.334)\frac{177-1}{177-5} \approx 0.318$$

c. Teste se é aceitável concluir que um aumento de 2% nas vendas origine um aumento de 0.5% no salário do CEO.

$$H_0: \beta_3 = 0.25, \qquad H_1: \beta_3 \neq 0.25$$

$$ET: T = \frac{b_3 - \beta_3^0}{se_{b_3}} \sim t_{(n-k)} \qquad \text{com } n=177 \text{ e } k=5$$

$$t_{obs} = \frac{0.196 - 0.25}{0.033} = -1.636$$

$$W = \{t: t < -1.96 \vee t > 1.96\} \quad \text{e portanto } t_{obs} \notin W$$

Não se rejeita H_0 num teste com 5% de significância, logo, com os dados de que dispomos, não temos razões para excluir a hipótese de que o verdadeiro valor do coeficiente (na população) seja 0.25.

d. Formalize o teste de hipóteses relacionado com a estimação presente no **Anexo 2** e retire a sua conclusão.

$$H_0: \beta_4 = \beta_5 = 0, \qquad H_1: \exists \beta_j \neq 0, \qquad j = 4{,}5$$

$$ET: F = \frac{R^2 - R_0^2}{1 - R^2} \times \frac{n-k}{m} \sim F_{(m, n-k)} \quad \text{com} \quad m{=}2,\ n{=}177 \text{ e } k{=}5$$

$$F_{obs} = \frac{0.334 - 0.290}{1 - 0.334} \times \frac{177 - 5}{2} \approx 5.682$$

(Alternativamente para quem calcule o valor com base em VR vem $F_{obs} = \frac{(45.909 - 43.081)/2}{0.250} \approx 5.656$).

Como $W = \{f: f > 3\}$, rejeita-se H_0. Sendo assim, com os dados de que dispomos, concluímos que as variáveis *exper* e *exper2* são, conjuntamente, significativas.

e. Qual o objectivo de incluir a variável $exper2$ no modelo, dado que já tínhamos $exper$? Apresente o valor a partir do qual a experiência como gestor é, de acordo com o modelo, negativa para o salário.

Ao incluir a variável $exper2$ o efeito da experiência do gestor no seu salário deixa de ser linear. Dado que as estimativas têm sinais opostos, o efeito final resulta de um equilíbrio entre um efeito positivo e um efeito negativo.

$$0{,}045 - (2 \times 0{,}001)exper \leq 0 \Leftrightarrow exper \geq 22.5$$

Neste caso, estima-se que em média, a partir do momento em que o gestor passe dos 22 anos e 6 meses de experiência, o efeito desta variável no seu salário comece a ser negativo.

f. Após a estimação do modelo proposto decidiu-se averiguar se era razoável admitir que os erros no modelo inicialmente proposto eram homocedásticos, tendo-se obtido o **Anexo 3**. Qual a variável dependente da regressão? Formalize e efectue o teste estatístico adequado, considerando uma dimensão de 5%.

$$\hat{u}_t^2 = \alpha_1 + \alpha_2 \widehat{lsal}_t + \alpha_3 \widehat{lsal}_t^2 + v_t$$

$$H_0: \alpha_2 = \alpha_3 = 0, \qquad H_1: \exists \alpha_j \neq 0, \qquad j = 2{,}3$$

$$ET: F = \frac{R^2 - R_0^2}{1 - R^2} \times \frac{n-k}{m} \sim F_{(m, n-k)} \qquad F_{obs} = 0.175, \ P_{obs} = 0.839$$

$$ou$$

$$\text{E.T.}: W = n \times R^2 \overset{\circ}{\sim} \chi^2_{(2)} \qquad W_{obs} = 0.354 \qquad p_{obs} = 0.838$$

Não encontrámos evidência estatística contra H_0, logo, com os dados de que dispomos, não temos razões para por em causa a hipótese de homocedasticidade.

Anexo 1 – Variável dependente *lsal*

Regression Statistics	
R Square	0.3336
Adjusted R Square	[redacted]
Standard Error	0.5005
Observations	177

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	4	21.566	5.391	21.525	0.000
Residual	172	43.081	0.250		
Total	176	64.646			

	Coefficients	*Standard Error*	*t Stat*	*P-value*
Intercept	4.90417	0.24397	20.458	0.000
lucros	0.00018	0.00012	1.566	0.119
lvendas	0.19636	0.03314	5.925	0.000
exper	0.04537	0.01422	3.191	0.002
exper2	-0.00123	0.00048	-2.572	0.011

Anexo 2 – Variável dependente *lsal*

Regression Statistics	
R Square	0.290
Adjusted R Square	0.282
Standard Error	0.514
Observations	177

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	2	18.737	9.369	35.507	0.000
Residual	174	45.909	0.264		
Total	176	64.646			

	Coefficients	*Standard Error*	*t Stat*	*P-value*
Intercept	5.14509	0.23473	21.919	0.000
lucros	0.00018	0.00012	1.483	0.140
lvendas	0.19370	0.03400	5.697	0.000

Anexo 3

Regression Statistics	
R Square	0.002
Adjusted R Square	-0.009
Standard Error	0.533
Observations	177

ANOVA

	Df	*SS*	*MS*	*F*	*Significance F*
Regression	2	0.099	0.050	0.175	0.839
Residual	174	49.378	0.284		
Total	176	49.477			

	Coefficients	*Standard Error*	*t Stat*	*P-value*
Intercept	-3.482	10.019	-0.348	0.729
Predicted lsal	1.174	3.020	0.389	0.698
Predicted lsal^2	-0.092	0.227	-0.405	0.686

Nota: Predicted lsal - valores ajustados de *lsal* na regressão do Anexo 1. Predicted lsat^2 – quadrado dos valores ajustados de *lsal* na regressão do anexo 1.