Instituto Superior de Economia e Gestão
UNIVERSIDADE TÉCNICA DE LISBOA

Estatística II – Licenciatura em Gestão
EN - 12 Junho 2012

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**
Quando num teste de hipóteses com $\alpha = 0.01$ se obtém um valor-p de 0.0025 rejeita-se H_0	X	
Quando, num modelo de regressão linear, se introduz uma variável explicativa adicional, pode acontecer que o valor do coeficiente de determinação diminua		X
Num teste de hipóteses, as alternativas definidas por H0 e por H1 devem ser disjuntas	X	
Sejam T_1 e T_2 dois estimadores de θ com $\mathrm{var}(T_1) < \mathrm{var}(T_2)$. Pode se garantir que T_1 tem um erro quadrático médio inferior ao de T_2		X
No output gerado pelo EXCEL quando se activa o procedimento de regressão para estimar o modelo $y_t = \beta_1 + \beta_2 x_{t2} + \beta_3 x_{t3} + u_t$ obtém-se uma estimativa da matriz das covariâncias dos estimadores dos parâmetros β_j ($j = 1,2,3$).		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. O lema de Neyman-Pearson é um resultado teórico importante no estudo
      - ☐ Do modelo de regressão linear
      - ☐ Do enviesamento de um estimador
      - X Dos testes de hipóteses
      - ☐ Da eficiência de um estimador

   b. Quando, depois de estimar o modelo $y_t = \beta_1 + \beta_2 x_{t2} + u_t$ se estima o modelo $\hat{u}_t^2 = \alpha_1 + \alpha_2 x_{t2} + \alpha_3 x_{t2}^2 + v_t$ é porque
      - ☐ Se suspeita de autocorrelação condicionada
      - X Se suspeita de heterocedasticidade condicionada
      - ☐ Se suspeita de alteração de estrutura
      - ☐ Se suspeita que a variável u_t não tem distribuição normal

   c. No modelo de regressão linear, $y_t = \beta_1 + \beta_2 x_{t2} + \cdots + \beta_k x_{tk} + u_t$, $t = 1,2,\cdots,n$, ao recorrer ao método dos mínimos quadrados para estimar os parâmetros β_j, $j = 1,2,\cdots,k$
      - ☐ Garante-se que não existem resíduos negativos
      - ☐ Garante-se que não existem resíduos positivos
      - ☐ Procura penalizar-se de igual forma todos resíduos
      - X Dá-se maior penalização aos grandes resíduos quando avaliados em valor absoluto

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

a. Defina o conceito de erro quadrático médio e refira a sua utilidade

Seja T um estimador de θ. $EQM(T) = E(T-\theta)^2$

O EQM serve para comparar a qualidade de estimadores que não sejam forçosamente centrados

b. Considere o MRL $y_t = \beta_1 + \beta_2 x_{t2} + \beta_3 x_{t3} + \beta_4 x_{t4} + u_t$ a verificar as hipóteses habituais (H1 a H6). Deduza a forma da regressão auxiliar que irá utilizar para testar $H_0: \begin{cases} \beta_2 = 3\beta_4 \\ \beta_3 = 1.5 \end{cases}$.

Dado H_0, vem $y_t = \beta_1 + 3\beta_4 x_{t2} + 1.5 x_{t3} + \beta_4 x_{t4} + u_t$, isto é, $y_t - 1.5 x_{t3} = \beta_1 + \beta_4 (3x_{t2} + x_{t4}) + u_t$

Assim a regressão auxiliar será $y_t^* = \alpha_1 + \alpha_2 x_t^* + v_t$ com $y_t^* = y_t - 1.5 x_{t3}$ e $x_t^* = 3x_{t2} + x_{t4}$.

Estatística II – Licenciatura em Gestão
EN - 12 Junho 2012

Parte Prática

Nome: ____ **Nº**________

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

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 folha em branco no fim do enunciado

1. Um radar está colocado à saída de um túnel para detectar a velocidade dos automóveis. Sempre que o radar capta velocidades acima dos 50km/h, o automóvel em questão é identificado para ser aplicada a respectiva multa. Contudo, o desempenho do radar tem vindo a piorar ao longo do tempo, pelo que a autarquia local decidiu analisar o problema. Sendo X a variável aleatória que representa o erro de medida na velocidade (isto é, a diferença entre a velocidade registada pelo radar e a verdadeira velocidade do automóvel), a autarquia está em condições de afirmar que este erro de medida tem distribuição normal com média nula e variância desconhecida e igual a θ, isto é, $X \sim N(0;\theta)$.

   a) Para analisar a performance do radar, a autarquia decidiu recolher uma amostra aleatória de n passagens de automóveis. Mostre que o estimador obtido pelo método dos momentos para a $\theta = \mathrm{var}(X)$ é dado por $\tilde{\theta} = \frac{1}{n}\sum_{i=1}^{n} X_i^2$. (Note que apenas necessita recorrer aos segundos momentos).

Recorrendo aos 2os momentos vem :

População: $E(X^2) = \sigma^2 + \mu^2 = \sigma^2 = \theta$

Amostra: $M_2 = \dfrac{\sum_{i=1}^{n} X_i^2}{n}$

Logo o estimador dos momentos será a solução de $\dfrac{\sum_{i=1}^{n} X_i^2}{n} = \theta$ isto é $\tilde{\theta} = \dfrac{\sum_{i=1}^{n} X_i^2}{n}$

b) Sabendo que se observou $s'^2_X = 4.485$ numa amostra casual com 41 observações, obtenha um intervalo de confiança a 95% para θ, variância da população.

Variável fulcral: $$Q = \frac{(n-1)S'^2}{\sigma^2} = \frac{(n-1)S'^2}{\theta} \sim \chi^2(n-1) \text{ com } n = 41$$

$q_1 = 24.433$; $q_2 = 59.342$ Percentis 0.025 e 0.975 de uma qui-quadrado com 40 g.l.

Intervalo de confiança

$\left(\frac{(n-1)\,s'^2}{q_2}; \frac{(n-1)\,s'^2}{q_1}\right)$ isto é $\left(\frac{40\times 4.485}{59.342}; \frac{40\times 4.485}{24.433}\right)$ ou seja (3.023; 7.343)

c) Devido ao mau comportamento do radar em algumas situações a autarquia resolveu estudar a sua troca por outro mais moderno. O vendedor afirmou, sem sombra de dúvida, que o novo radar tem um erro de medida, Y, com distribuição normal de média nula e variância muito inferior $Y \sim N(0, \sigma_Y^2)$. Ao analisar o desempenho deste novo radar, também com base numa amostra de 41 observações, a autarquia obteve $s'^2_Y = 0.233$. Na amostra de 41 observações com o radar antigo tinha-se obtido $s'^2_X = 4.485$. Teste para $\alpha = 0.05$ a hipótese $H_0 : \frac{\sigma_X^2}{\sigma_Y^2} = 1$ contra $H_1 : \frac{\sigma_X^2}{\sigma_Y^2} > 1$ e comente a afirmação do vendedor.

Universos normais. Teste de $H_0 : \frac{\sigma_X^2}{\sigma_Y^2} = 1$ contra $H_1 : \frac{\sigma_X^2}{\sigma_Y^2} > 1$.

Estatística de teste: $F = \frac{S'^2_X}{S'^2_Y} \sim F(m-1, n-1)$ sendo $m = n = 41$

$W_F = \{f : f > f_\alpha\}$ com $f_\alpha = 1.69$ → *F-Snedecor* com (40;40) g.l.

$f_{obs} = \frac{s'^2_X}{s'^2_Y} = \frac{4.485}{0.233} = 19.25 \in W_F$. Logo rejeita-se H_0, isto é, o vendedor tem razão: O novo radar tem uma variância inferior ao radar instalado.

2. Um analista de um departamento de gestão do risco de crédito de um Banco pretende saber se existe independência entre a taxa de recuperação de um crédito em que houve incumprimento da contraparte e a colateralização desse crédito (existência de um bem que sirva de garantia ao crédito). Para esse fim, recolheu uma amostra casual de 600 créditos em que houve incumprimento da contraparte e contou quantos casos observou em cada situação.

	recuperação nula	recuperação parcial	recuperação total
Com colateral	55	100	130
Sem colateral	200	80	35

O que pode o analista concluir?

Teste de independência: $H_0: p_{ij} = p_{i\bullet} \times p_{\bullet j}$, $i = 1,2$, $j = 1,2,3$ contra $H_1: \exists (i,j): p_{ij} \neq p_{i\bullet} \times p_{\bullet j}$

Estatística de teste: $Q = \sum_{i=1}^{2} \sum_{j=1}^{3} \frac{\left(N_{ij} - fe_{ij}\right)^2}{fe_{ij}} \overset{a}{\sim} \chi^2_{(1\times 2)}$ com $fe_{ij} = \frac{N_{i\bullet} \times N_{\bullet j}}{n}$

Frequências esperadas sob H_0

	nula	parcial	total
com colateral	121.13	85.50	78.38
sem colateral	133.88	94.50	86.63

$q_{obs} = 138.22$ valor-*p* < 0.001 (9.70E-31)

Rejeita-se a hipótese nula de independência entre a recuperação de crédito e a existência de colateral.

3. Num estudo econométrico sobre fiscalidade das empresas portuguesas, foi utilizada uma amostra casual de 142 empresas tendo-se observado as seguintes variáveis: IMPOS = Impostos (em unidades monetárias); VOL_NEG = Volume de negócios (em unidades monetárias); EMPREG = Número de empregados; e AMORT = Amortizações (em unidades monetárias). Com base na ferramenta "Regression" do programa EXCEL, obtiveram-se os outputs dos modelos A, B e C. Face ao exposto:

   a) Interprete as estimativas obtidas para os coeficientes de regressão de ln(VOL_NEG) e EMPREG no modelo (A).

$b_{\ln(VOL\_NEG)} = b_2 = 1.0423$ *Ceteris paribus*, um aumento de 1% no volume de negócios provoca um acréscimo esperado de aproximadamente 1.04% nos impostos pagos pelas empresas portuguesas.

$b_{EMPREG} = b_3 = -0.0000776$ *Ceteris paribus*, um aumento dos recursos humanos em 1 unidade origina um decréscimo de aproximadamente 100×0.0000776=0.00776% nos impostos a pagar pelas empresas. No caso do aumento dos recursos humanos ser de 100 unidades, o decréscimo nos impostos seria de aproximadamente 0.776%.

b) Ainda com base nos resultados obtidos para o modelo (A), teste a significância individual das amortizações e a significância global da regressão. Complete também o quadro ANOVA indicando o valor que deveria figurar na linha "regression", coluna "SS".

$H_0 : \beta_4 = 0 \quad \text{contra} \quad H_1 : \beta_4 \neq 0$

Estatística de teste: $T = \dfrac{b_4 - \beta_4}{s_{b_4}} \sim t_{(138)}$

$t_{obs} = 1.4961$ e valor-*p*=0.1369 (ver output) logo não se rejeita a hipótese nula. A variável amortizações não tem relevância estatística para explicar o logaritmo dos impostos.

$H_0 : \beta_2 = \beta_3 = \beta_4 = 0 \quad \text{contra} \quad H_1 : \exists \beta_j \neq 0,\ j = 2,3,4$

Estatística de teste: $F = \dfrac{R^2 / (k-1)}{(1-R^2)/(n-k)} \sim F(k-1, n-k)$ ou $F = \dfrac{R^2 / 3}{(1-R^2)/138} \sim F(3,138)$

$F_{obs} = 37.9216$ e valor-p=6.14E-18 (ver output) logo, deve rejeitar-se a hipótese nula e concluir que a regressão é globalmente significativa.

O valor pedido é dado por $3 \times 21.8611 = 65.5833$

c) Teste a nulidade conjunta dos coeficientes β_3 e β_4 no Modelo (A). Que conclusão tira no que se refere aos respectivos regressores?

$H_0 : \beta_3 = \beta_4 = 0 \quad \text{contra} \quad H_1 : \exists \beta_j \neq 0,\ j = 3,4$

Estatística de teste: $F = \dfrac{(R^2 - R_0^2)/m}{(1-R^2)/(n-k)} \sim F(m, n-k)$ **ou** $F = \dfrac{(VR_0 - VR_1)/m}{VR_1/(n-k)} \sim F(m, n-k)$

em que o índice 0 se refere ao Modelo (B) e o índice 1 (ou sem índice) ao Modelo (A)

$$F_{obs} = \frac{(0.4519 - 0.4408)/2}{(1-0.4519)/(142-4)} = 1.397 < F_{0.05} \approx 3.06$$

Logo, não se rejeita a hipótese nula e conclui-se que se podem eliminar simultaneamente os regressores EMPREG e AMORT.

Alternativamente, $valor - p = P(F > 1.397) = 0.2507$ e mesma conclusão.

d) Com base no **Modelo (B)** e usando um teste estatístico adequado, comente a seguinte afirmação: "*Ceteris paribus*, um aumento de 10% no volume de negócios, leva a um aumento de aproximadamente 12% nos impostos pagos pelas empresas portuguesas".

$$H_0: \beta_2 = \frac{12\%}{10\%} = 1.2 \quad \text{contra} \quad H_1: \beta_2 \neq 1.2$$

Estatística de teste: $T = \dfrac{b_2 - \beta_2}{s_{b_2}} \sim t_{(138)}$

$$|t_{obs}| = \left|\frac{b_2 - \beta_2}{s_{b_2}}\right| = \left|\frac{1.0884 - 1.2}{0.1036}\right| = |-1.077| < t_{0.05/2}(142-4) \approx 1.977$$

Não se rejeita a hipótese nula, isto é, a afirmação tem fundamentação estatística.

Alternativamente,

$Valor - p = 2P(T > |t_{obs}| \mid H_o) = 2P(T > 1.077) = 0.2833$ e mesma conclusão.

e) Para que serve a regressão auxiliar dada pelo modelo (C)? Efectue o teste estatístico adequado e conclua.

A regressão auxiliar serve para testar a hipótese de homocedasticidade condicionada da variável residual. Para testar esta hipótese recorre-se à regressão auxiliar dada pelo Modelo C e testa-se $H_0: \beta_2 = \beta_3 = 0$ contra $H_1: \exists \beta_j \neq 0,\ j = 2,3$ utilizando o teste de White.

Estatística de teste: $W = nR^2 \overset{a}{\sim} \chi^2_{(2)}$ e m que R^2 é o coef. de determinação da regressão auxiliar.

$W_{obs} = 142 \times 0.0432 = 6.1344 > \chi^2_{0.05} = 5.99147$ **ou** valor-p=0.0466

Considerando uma dimensão de teste de 5%, rejeita-se a hipótese nula e conclui-se que as variáveis residuais não são homoscedásticas. A conclusão seria diferente se se considerasse uma dimensão de 1% ou até de 4%.

Modelo (A): $\boldsymbol{ln(IMPOS_t) = \beta_1 + \beta_2 ln(VOL\_NEG_t) + \beta_3 EMPREG_t + \beta_4 AMORT_t + u_t}$

Regression Statistics	
Multiple R	0.6722
R Square	0.4519
Adjusted R Square	0.4400
Standard Error	0.7593
Observations	142

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	3	**?**	21.8611	37.9216	6.14E-18
Residual	138	79.5544	0.5765		
Total	141				

	Coefficients	*Standard Error*	*t Stat*	*P-value*
Intercept	-3.0367	0.8487	-3.5780	0.0005
ln(VOL_NEG)	1.0423	0.1273	8.1890	1.58E-13
EMPREG	-7.76E-05	0.0005	-0.1486	0.8821
AMORT	2.97E-08	1.98E-08	1.4961	0.1369

Modelo (B): $\boldsymbol{ln(IMPOS_t) = \beta_1 + \beta_2 ln(VOL\_NEG_t) + u_t}$

Regression Statistics	
Multiple R	0.6640
R Square	0.44080
Adjusted R Square	0.43680
Standard Error	0.7614
Observations	142

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	63.9811	63.9811	110.3712	2.14E-19
Residual	140	81.1566	0.5797		
Total	141	145.1377			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	-3.3392	0.7127	-4.6851	6.56E-06	-4.7483	-1.9301
ln(VOL_NEG)	1.0884	0.1036	10.5058	2.14E-19	0.8835	1.2932

Modelo (C): $\boldsymbol{\hat{u}_t^2 = \beta_1 + \beta_2 ln(VOL\_NEG_t) + \beta_3[ln(VOL\_NEG_t)]^2 + v_t}$**,** em que $\hat{u}_t^2$ representa o quadrado dos resíduos da estimação do modelo (B).

Regression Statistics	
Multiple R	0.2079
R Square	0.0432
Adjusted R Square	0.0294
Standard Error	0.7201
Observations	142

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	2	3.2559	1.6280	3.1391	0.0464
Residual	139	72.0863	0.5186		
Total	141	75.3422			

	Coefficients	*Standard Error*	*t Stat*	*P-value*
Interceptar	16.5121	6.6134	2.4967	0.0137
ln(VOL_NEG)	-4.4962	1.8945	-2.3733	0.0190
(ln(VOL_NEG))^2	0.3141	0.1349	2.3285	0.0213