Instituto Superior de Economia e Gestão
Universidade de Lisboa
Estatística II – Licenciatura em Gestão
Exame de Época de Recurso – 29 de Janeiro de 2014

Parte Teórica

Nome:____**nº processo:**_______________

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
Um estimador centrado com variância a tender para zero quando a dimensão da amostra tende para infinito é consistente.	X	
Num teste de dimensão 5 %, a probabilidade de não rejeitar a hipótese nula quando ela é verdadeira é igual a 0.95.	X	
Se (a, b) é um intervalo de confiança a 90% para o parâmetro θ, então $P(a < \theta < b) = 0.90$.		X
Considere o MRL $y_t = \beta_1 + \beta_2 x_t + u_t$. No teste de dimensão 5%, $H_0: \beta_2 = 0$ contra $H_1: \beta_2 \neq 0$, a conclusão "β_2 é estatisticamente significativo a 5%" é equivalente a "valor-p < 0.05".	X	
Os testes de alteração da estrutura não podem ser realizados recorrendo a variáveis artificiais		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) Considere um teste para a média de uma população normal de variância conhecida: $H_0: \mu = 1$ contra $H_1: \mu = 2$, com região de rejeição dada por $W_{\bar{x}} = \{\bar{x}: \bar{x} > 1.023\}$. A probabilidade de cometer o erro de 2ª espécie é dada por:

☐ $P(\bar{X} \leq 1.023 | \mu = 1)$
☐ $P(\bar{X} > 1.023 | \mu = 1)$
X $P(\bar{X} \leq 1.023 | \mu = 2)$
☐ $P(\bar{X} > 1.023 | \mu = 2)$

b) Admita que no MRL, $y_t = \beta_1 + \beta_2 x_{t2} + \beta_3 x_{t3} + u_t$, se pretende testar H_0: $\beta_2 = -\beta_3$ contra H_1: $\beta_2 \neq -\beta_3$. Supondo que se verificam as hipóteses habituais, o teste pode ser realizado utilizando a seguinte regressão auxiliar:

☐ $y_t = \beta_1 + \beta_2(-x_{t2}) + \beta_3 x_{t3} + u_t$.
X $y_t = \beta_1 + \beta_3(x_{t3} - x_{t2}) + u_t$.
☐ $y_t = \beta_1 + \beta_2(x_{t2} + x_{t3}) + u_t$.
☐ $y_t = \beta_1 + u_t$.

c) Considere o seguinte quadro ANOVA decorrente da estimação do modelo $y_t = \beta_1 + \beta_2 x_{t2} + \beta_3 x_{t3} + u_t$,

ANOVA

	df	*SS*	*MS*	*F*
Regression	2	17.1069	8.5535	33.0760
Residual	115	29.7391	0.2586	
Total	117	46.8460		

O coeficiente de determinação, R^2, é igual a:

☐ 0.2586 ☐ 0.6348 X 0.3652 ☐ a informação dada não permite obter R^2

3. Perguntas de desenvolvimento [2.25 valores: a) 1 valor; b) 1.25 valores]

a)Defina os conceitos de hipótese estatística e de hipótese paramétrica.

Uma hipótese estatística é qualquer conjectura sobre aspectos desconhecidos da função de distribuição de determinada população. Quando a forma da função de distribuição ou da função densidade é conhecida e a conjectura diz respeito apenas ao parâmetro (escalar ou vector), tem-se uma hipótese paramétrica.

b) Para analisar as despesas das famílias com "produtos verdes" (alimentos biológicos, produtos economizadores de energia, etc...) foi especificado o seguinte modelo:

$$ldesbio_t = \beta_1 + \beta_2 lrend_t + \beta_3 idade_t + u_t,$$

onde *ldesbio* representa o logaritmo natural das despesas da família com "produtos verdes", *lrend* é o logaritmo natural do rendimento médio da família e *idade* é a idade do chefe de família.

Reespecifique o modelo de forma que **simultaneamente:** i) a elasticidade dos gastos com produtos verdes relativamente ao rendimento possa variar de acordo com a zona de residência (urbana ou rural) e ii) o termo independente possa variar com o nível de escolaridade do chefe de família (básico, secundário ou superior). **Nota**: explicite claramente a(s) variável(is) que necessita empregar.

O novo modelo é, por exemplo,

$$ldesbio_t = \beta_1 + \delta_1 bas_t + \delta_2 sec_t + \beta_2 lrend_t + \delta_3 urbana_t \times lrend_t + \beta_3 idade_t + v_t,$$

Onde *bas* é uma variável artificial que assume o valor 1 se o chefe de família tem o nível básico de escolaridade e o valor 0, caso contrário; *sec* é uma variável artificial que assume o valor 1 se o chefe de família tem o nível secundário de escolaridade e o valor 0, caso contrário; *urbana* é uma variável artificial que assume o valor 1 se família reside numa zona urbana e o valor 0, caso contrário.

Instituto Superior de Economia e Gestão
Universidade de Lisboa
Estatística II – Licenciatura em Gestão
Exame de Época de Recurso – 29 de Janeiro de 2014

Parte Prática

Nome:____**nº processo:**_______________

Espaço reservado para classificações				
1a. (15) ____	2a. (20) ____	3a. (15) ____	3c. (20) ____	**Teórica:**
1b. (20) ____	2b. (15) ____	3b. (15) ____	3d. (20) ____	**Prática:**

Em todos os testes de hipóteses que realizar, 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. Considere sempre uma dimensão de teste de 0.05. Se necessitar de espaço para continuar qualquer resposta dispõe de uma página em branco no fim do enunciado.

1. Um inquérito a 500 portugueses com mais de 18 anos sobre a utilização da internet na compra de bens ou serviços permitiu resumir os resultados no seguinte quadro:

Idade \ **Utilização da internet**	compra *online*	não compra *online*
18 - 30	42	108
30 - 50	28	132
>50	20	170

a) Com base num teste adequado ao nível de 5 % diga se concorda com a seguinte afirmação "No máximo, 16 % dos portugueses efectuam compras de bens ou serviços *online*".

Seja $X = 1$, se o indivíduo compra bens ou serviços *online* ; $X \sim B(1,\theta)$ com θ, proporção de portugueses que compram bens ou serviços *online*; $n = 500$; $\bar{x} = \frac{90}{500} = 0.18$.

Pretende-se testar: $H_0: \theta \leq 0.16$ contra $H_1: \theta > 0.16$

Estatística de teste sob H_0: $\frac{\bar{X}-\theta_0}{\sqrt{(\theta_0(1-\theta_0))/n}} = \frac{\bar{X}-0.16}{\sqrt{(0.16\times0.84)/n}} \dot{\sim} N(0,1)$; $z_{obs} = \frac{0.18-0.16}{\sqrt{(0.16\times0.84)/500}} = 1.2195$.

Região de rejeição a 5%: $W = \{z: z > 1.645\}$. Como $z_{obs} \notin W$, não se rejeita H_0 ao nível de 5%: a evidência estatística leva-nos a concordar com a afirmação.

b) Ao nível de 5 %, pode afirmar-se que a utilização da internet para compras de bens ou serviços é independente da idade?

Teste de independência: $H_0: p_{ij=}p_{i0} \times p_{0j}$ $(i = 1,2,3 ; j = 1,2)$ contra $H_1: \exists\, i,j: p_{ij\neq}p_{i0} \times p_{0j}$.

Estatística de teste sob H_0: $Q = \sum_i \sum_j \frac{(N_{ij}-fe_{ij})^2}{fe_{ij}} \dot{\sim} \chi^2(2)$; $W_{5\%} = \{q: q > 5.991\}$.

$$q_{obs} = \frac{(42-27)^2}{27} + \frac{(108-123)^2}{123} + \cdots + \frac{(170-155.8)^2}{155.8} = 17.3798.$$

Como $q_{obs} \in W_{5\%}$, rejeita-se H_0 a 5%: não é de admitir independência entre utilização da internet para compras de bens ou serviços e idade.

2. Seja X uma variável aleatória de média $\theta+4$, variância θ^2 e função densidade

$$f(x\,|\,\theta)=\frac{1}{\theta}e^{-\frac{x-4}{\theta}}\ ,\quad x>4\,,\theta>0\,.$$

a) Dada a amostra casual $(X_1,X_2,\ldots,X_n)$, mostre que o estimador de máxima verosimilhança para θ é $\hat{\theta}=\bar{X}-4$. Considere apenas a condição de primeira ordem.

$\hat{\theta}$: max $L(\theta)$, com $L(\theta)=\prod_{i=1}^{n}f\,(x_i|\theta)=\frac{1}{\theta^n}e^{-\frac{\sum(x_i-4)}{\theta}}$

$l(\theta)=$log$L(\theta)=-nlog(\theta)-\frac{\sum(x_i-4)}{\theta}$

$$\frac{dl(\theta)}{d\theta}=-\frac{n}{\theta}+\frac{\sum(x_i-4)}{\theta^2}$$

$\frac{dl(\theta)}{d\theta}=0\Leftrightarrow -n\theta+\sum(x_i-4)=0\Leftrightarrow\hat{\theta}=\bar{X}-4.$

Nota: verifica-se: $\frac{d^2l(\theta)}{d\theta^2}<0.$

b) Analise o enviesamento e a consistência de $\hat{\theta}$.

- $\hat{\theta}$ é um estimador centrado para o parâmetro θ se $E(\hat{\theta})=\theta$

$E(\hat{\theta})=E(\bar{X}-4)=E(\bar{X})-4=E(X)-4=(\theta+4)-4=\theta$. Como $E(\hat{\theta})=\theta$, $\hat{\theta}$ é centrado.

- Se $\lim_{n\to\infty}E(\hat{\theta})=\theta$ e $\lim_{n\to\infty}var(\hat{\theta})=0$, $\hat{\theta}$ é um estimador consistente para o parâmetro θ.

Como $var(\hat{\theta})=var(\bar{X}-4)=var(\bar{X})=\frac{var(X)}{n}=\frac{\theta^2}{n}$, obtém-se:

$$\lim_{n\to\infty}E(\hat{\theta})=\lim_{n\to\infty}\theta=\theta \text{ e } \lim_{n\to\infty}var(\hat{\theta})=\lim_{n\to\infty}\frac{\theta^2}{n}=0$$

e portanto, $\hat{\theta}$ é consistente.

3. Para analisar o comportamento dos salários (*sal*) dos CEO das empresas de determinado país, foi proposto o seguinte modelo:

$$lsal_t=\beta_1+\beta_2 fin_t+\beta_3 lvendas_t+\beta_4 ant_t+\beta_5 ant_t^2+u_t$$

onde *lsal* representa o logaritmo do salário do CEO, *lvendas* representa o logaritmo das vendas da empresa onde trabalha o CEO, *ant* é o número de anos como CEO da empresa e *fin* é uma variável artificial que assume o valor 1 se o indivíduo é CEO de uma empresa financeira. Os resultados obtidos com base numa amostra de 177 observações encontram-se no Anexo onde, em todos os modelos, a variável dependente é *lsal* e, no modelo 1, $ant2=ant^2$.

Nota: Todas as perguntas têm por base o modelo 1 do Anexo.

a) Interprete a estimativa do parâmetro β_2 e teste a significância estatística de β_2.

$b_2=-0.0353$: estima-se que um CEO de uma empresa financeira ganha menos, em média, cerca de 3.53% do que um CEO de uma empresa não financeira, para valores iguais das restantes variáveis.

H_0: $\beta_2=0$ contra $\beta_2\neq 0$; sob H_0, $T=\frac{b_2}{s_{b_2}}\sim t(n-k)=t(172)$. Como valor-p = 0.6458, não se rejeita H_0 : β_2 não é estatisticamente significativo.

b) Afirma-se que a elasticidade do salário em relação às vendas se situa entre 0.1850 e 0.2725, com determinado nível de confiança. Determine, justificando, esse nível de confiança.

Intervalo de confiança a $(1-\alpha)\times 100\%$ para a elasticidade β_3 é $(b_3 \mp t\alpha_{/2} \times s_{b_3})$ obtido com base na variável fulcral $T = \frac{b_3-\beta_3}{s_{b_3}} \sim t(n-k) = t(172)$.

Como a amplitude do intervalo de confiança é dada por $\Delta = 2 \times t\alpha_{/2} \times s_{b_3} = 0.2725 - 0.1850 = 0.0875$, obtém-se $t\alpha_{/2} = \frac{0.0875}{2\times s_{b_3}} = 1.645$. Como $\mathrm{P}\left[T > t\alpha_{/2}\right] = \mathrm{P}\,[T > 1.645] = 0.05$, obtém-se $\frac{\alpha}{2} = 0.05 \Leftrightarrow (1-\alpha) = 0.90$.

Conclui-se que (0.1850, 0.2725) é um intervalo de confiança para β_3 a 90 %.

c) Teste, ao nível de 5%, a validade da seguinte afirmação: “Os anos como CEO da empresa são relevantes para explicar o salário”.

Pretende−se testar: $H_0: \beta_4 = \beta_5 = 0$ contra $H_1: \beta_4 \neq 0$ ou $\beta_5 \neq 0$.

sob H_0, $F = \frac{(VR_0 - VR_1)/2}{VR_1/(n-k)} \sim F(2{,}172)$; $W_{5\%} = \{f: f > 3.0\}$. Como $f_{obs} = \frac{(46.4268-43.6406)/2}{43.6406/172} = 5.49$, rejeita-se H_0 a 5%: β_4 e β_5 são, em conjunto estatisticamente significativos: é de admitir que a antiguidade contribui para explicar o comportamento dos salários dos CEO.

d) Suspeitando-se de que poderiam não ser válidas algumas das hipóteses do MRL, estimou-se a seguinte regressão onde $\hat{u}$ e $\widehat{lsal}$ são, respectivamente, os resíduos e os valores ajustados do modelo inicial:

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

Indique o objectivo dessa estimação. Sabendo que se obteve $R^2 = 0.00068$, tire as conclusões que achar adequadas, justificando devidamente a sua resposta.

Pretende-se testar a presença de heterocedasticidade condicionada nos erros do modelo 1, utilizando o teste simplificado de White.

$H_0: \sigma_1^2 = \sigma_2^2 = \cdots = \sigma_n^2$ contra $H_1: \sigma_t^2 \neq \sigma_s^2$ para algum $t \neq s$.

Sob H_0, $W_s = nR^2 \dot{\sim} \chi^2(2)$; $W_{5\%} = \{q: q > 5.991\}$. Como $q_{obs} = 177 \times 0.00068 = 0.12036 \notin W_{5\%}$, a hipótese nula não é rejeitada a 5 %: não há evidência de heterocedasticidade condicionada nos erros. Supondo que se verificam as restantes hipóteses, o estimador dos mínimos quadrados é BLUE e a inferência habitual é válida.

Anexo

Modelo 1

Regression Statistics	
Multiple R	0.5700
R Square	0.3249
Adjusted R Square	0.3092
Standard Error	0.5037
Observations	177

ANOVA

	df	*SS*	*MS*	*F*
Regression	4	21.0056	5.2514	20.6973
Residual	172	43.6406	0.2537	
Total	176	64.6462		

	Coefficients	*Standard Error*	*t Stat*	*P-value*
Intercept	4.7282	0.2105	22.4664	0.0000
fin	-0.0353	0.0767	-0.4604	0.6458
lvendas	0.2288	0.0266	8.5931	0.0000
ant	0.0452	0.0143	3.1570	0.0019
ant2	-0.0012	0.0005	-2.5533	0.0115

Modelo 2

Regression Statistics	
Multiple R	0.5309
R Square	0.2818
Adjusted R Square	0.2736
Standard Error	0.5165
Observations	177

ANOVA

	df	*SS*	*MS*	*F*
Regression	2	18.2194	9.1097	34.1416
Residual	174	46.4268	0.2668	
Total	176	64.6462		

	Coefficients	*Standard Error*	*t Stat*	*P-value*
Intercept	4.9732	0.2019	24.6260	0.0000
fin	-0.0379	0.0781	-0.4859	0.6276
lvendas	0.2254	0.0273	8.2608	0.0000