Instituto Superior de Economia e Gestão
UNIVERSIDADE TÉCNICA DE LISBOA

Estatística II – Licenciatura em Gestão

ER – 2 Julho 2013

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 do χ^2 à bondade do ajustamento a região de rejeição pode ser bilateral		
Quando num teste de hipóteses com $\alpha = 0.03$ se obtém um valor-p de 0.047 rejeita-se H_0		
Seja X uma população com distribuição exponencial. A conjectura $P(X < 2) = 0.4$ pode ser testada recorrendo a um teste de hipóteses paramétricas.		
Na estimação do MRL $y_t = \beta_1 + \beta_2 x_{t2} + \beta_3 x_{t3} + u_t$, $t = 1,2,...,105$, a verificar as hipóteses habituais obteve-se $R^2 = 0.971$. Pode-se então garantir que se rejeita, para os níveis de significância habituais, H_0 no teste $H_0 : \beta_2 = 0$ contra $H_1 : \beta_2 \neq 0$.		
Quando uma população não é normal a média da amostra é um estimador enviesado da média da população, se esta existir		

2. **Perguntas de resposta múltipla** (**2.25 valores**) - Para cada pergunta escolha **a** alternativa correcta. Uma resposta certa vale 0.75 valores e uma resposta errada penaliza em 0.25 valores.

   a. Considere um teste de independência χ^2 construído com base numa tabela de contingência. Assinale a alternativa correcta
      - ☐ O número de elementos observados em cada célula tem de ser ≥ 5
      - ☐ A hipótese H_0 traduz a independência entre os 2 factores em análise
      - ☐ Os graus de liberdade da qui-quadrado dependem de n , dimensão da amostra.
      - ☐ A região de rejeição é bilateral

   b. Seja o modelo de regressão linear $y_t = \beta_1 + \beta_2 x_{t2} + \beta_3 x_{t3} + u_t$ com $t = 1,2,...,n$. Quando se refere que o modelo não sofre de autocorrelação está-se a dizer que, para $t, s = 1,2,...,n$,
      - ☐ $\mathrm{cov}(x_{t2}, x_{s3} \mid X) = 0$
      - ☐ $\mathrm{cov}(u_t, u_s \mid X) = \sigma^2$ para $t \neq s$ e $\mathrm{cov}(u_t, u_s \mid X) = 0$ para $t = s$
      - ☐ $\mathrm{cov}(u_t, u_s \mid X) = 0$ para $t \neq s$
      - ☐ Todas as afirmações anteriores são falsas

   c. No modelo de regressão linear $y_t = \beta_1 + \beta_2 x_{t2} + \beta_3 x_{t3} + \beta_4 x_{t4} + \beta_5 x_{t5} + u_t$ com $t = 1,2,...,n$ pretende-se testar $H_0 : \beta_2 + \beta_3 = 0 \wedge \beta_4 = \beta_3$ contra $H_1 : H_0$ falsa . Para tal recorre-se à regressão auxiliar dada por
      - ☐ $y_t = \beta_1 + \beta_3(-x_{t2} + x_{t3} + x_{t4}) + \beta_5 x_{t5} + u_t$
      - ☐ $y_t = \beta_1 + \beta_3(x_{t2} + x_{t3} + x_{t4}) + \beta_5 x_{t5} + u_t$
      - ☐ $y_t = \beta_1 + \beta_3 x_{t3} + \beta_5 x_{t5} + u_t$
      - ☐ $y_t - x_{t2} = \beta_1 + \beta_3 x_{t3} + \beta_5 x_{t5} + u_t$

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

a. Defina o conceito de amostra emparelhada e explique o seu interesse.

b. Considere o MRL, $y_t = \beta_1 + \beta_2 x_t + u_t$, $t = 1, 2, \cdots, n$, onde $u_t = v_t \sqrt{x_t}$, com $E(v_t) = 0$, $\mathrm{var}(v_t) = \sigma^2$ e a variável v independente da variável x . Mostre que a hipótese H2, $E(u_t \mid X) = 0$, (exogeneidade condicionada) se encontra verificada e que a hipótese H4, $\mathrm{var}(u_t \mid X) = \sigma^2$, (homocedasticidade condicionada) se encontra violada.

Instituto Superior de Economia e Gestão
UNIVERSIDADE TÉCNICA DE LISBOA

Estatística II – Licenciatura em Gestão

ER – 3 Julho 2013

Parte prática

Nome: ______________________________ **Nº**__________

Espaço reservado para classificações				
1a. (15)	2a. (15)	4a. (15)	4d. (15)	**T:**
1b. (10)	2b.(15)	4b. (10)	4e. (15)	**P:** _____
	3.(15)	4c. (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, antes do anexo. Pode arrancar a folha de anexo se lhe der mais jeito

1. Seja X uma população com $E(X)=\frac{2}{1+\theta}$ e $Var(X)=\frac{2\theta(1-\theta)}{(1+\theta)^2}$, $\theta>0$ da qual se recolheu uma amostra casual de dimensão $n>5$.
   a. Para estimar $E(X)$ foi proposto o estimador $T=\frac{1}{8}(2X_1+5X_2+kX_5+4X_n)$, sendo k uma constante desconhecida. Determine k de forma a garantir que T é estimador centrado para a média da população. Obtenha também a variância do estimador T como função de k e de θ.

b. Estime θ pelo método dos momentos.

2. Uma máquina de café está afinada para encher um copo com 4 cl de café. A equipa de manutenção suspeita que a máquina está a fornecer menos quantidade de café do que aquela que estava prevista. Como o arranjo da máquina envolve despesas de algum montante a equipa de manutenção decidiu recolher uma amostra casual de 16 cafés tendo obtido $\sum_{i=1}^{16} x_i = 60.0,\ s'^2 = 0.16$. Assuma que a quantidade de café por copo segue uma distribuição normal.

a. Efectue um teste ($\alpha = 0.05$) que permita saber se a máquina deve ser sujeita a manutenção e conclua.

b. Com base na amostra observada, construiu-se, pelo processo habitual, o seguinte intervalo de confiança para a variância do conteúdo do copo (0.096; 0.331). Qual o grau de confiança associado ao intervalo?

3. Para melhor identificar o seu mercado alvo a Alfa Romeo conduziu um estudo de mercado. Uma amostra aleatória com 300 observações foi recolhida e cada pessoa seleccionada foi sujeita a um teste de condução findo o qual se classificou a sua atitude ao volante (Defensiva, Agressiva ou Equilibrada). Também se inquiriu, para cada pessoa, o seu modelo de Alfa Romeo preferido de entre 2 alternativas.

	Defensiva	Agressiva	Equilibrada	Total
Mito	60	10	30	100
Giulietta	60	60	80	200
Total	120	70	110	300

Teste (significância de 5%) se existe independência entre a atitude ao volante e o modelo de Alfa Romeo preferido.

4. Para analisar se o valor médio das rendas numa cidade é influenciado pela existência de uma universidade considerou-se o seguinte modelo:

$$\log(\text{rendas}_t)=\beta_1 + \beta_2 \log(\text{pop}_t) + \beta_3 \log(\text{rendmedio}_t) + \beta_4\, \text{univ}_t + \beta_5\, \text{propalug}_t + u_t \qquad t = 1, 2, \cdots, n$$

onde *log(rendas)* é o logaritmo do valor médio das rendas (em euros) na cidade t , *log(pop)* é o logaritmo da população da cidade t , *log(rendmedio)* é o logaritmo do rendimento médio em euros, *univ* variável binária que assume o valor 1 se existe uma universidade na cidade e *propalug* é a proporção de habitações alugadas no total de habitações ocupadas. No **anexo** são disponibilizados os resultados de algumas estimações. Em todas elas a variável dependente é o logaritmo do valor médio das rendas.

a. Interprete as estimativas obtidas para os coeficientes β_2 e β_4. Recorrendo ao valor-p o que pode dizer quanto à significância individual do regressor associado com β_2 ($\alpha = 0.10$)?

b. Interprete o valor obtido para o coeficiente de determinação e teste a significância global da regressão.

c. Teste a 5% se é razoável admitir que um acréscimo de 5% no rendimento médio origine um aumento de 4% no valor médio das rendas de determinada cidade.

d. Construa um intervalo de previsão a 90% para o valor médio das rendas numa cidade particular, a cidade A que é uma cidade universitária com 50500 habitantes. Considere ainda que, das habitações ocupadas, metade possui um contrato de aluguer e que o rendimento médio na cidade A é de 19 mil euros.

e. Após a estimação do modelo proposto decidiu-se averiguar se era razoável admitir que os erros no modelo inicialmente proposto eram homocedásticos. Apresente uma regressão auxiliar adequada para o efeito. Assuma que se estimou a regressão auxiliar que apresentou tendo-se obtido $R^2 = 0.424$. Efectue o teste considerando uma dimensão de 5%.

ANEXO

Modelo 1

SUMMARY
OUTPUT

Regression Statistics	
Multiple R	0.9179
R Square	0.8426
Adjusted R Square	0.8375
Standard Error	0.1341
Observations	128

ANOVA

	df	*SS*	*MS*	*F*
Regression	4	11.8456	2.9614	164.6359
Residual	123	2.21248	0.0180	
Total	127	14.0581		

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	-2.2166	0.4671	-4.7456	5.6664E-06	-3.1411	-1.2920
log(pop)	-0.0279	0.0203	-1.3746	0.1718	-0.0682	0.0123
log(rendmedio)	0.7996	0.0431	18.5523	1.8909E-37	0.7143	0.8849
univ	0.1492	0.0313	4.7601	5.3338E-06	0.0872	0.2112
propalug	0.0075	0.0014	5.5550	1.6345E-07	0.0049	0.0102

Modelo 2

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.9179
R Square	0.8426
Adjusted R Square	0.8375
Standard Error	0.1341
Observations	128

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	4	11.8457	2.9614	164.6359	2.1644E-48
Residual	123	2.2125	0.0180		
Total	127	14.0581			

	Coefficients	*Standard Error*	*t Stat*	*P-value*
Intercept	5.885	0.0269	219.1421	2.560E-161
log(pop)-log(50500)	-0.0279	0.0203	-1.3746	0.1718
log(rendmedio)-log(19000)	0.7996	0.0431	18.5523	1.891E-37
univ-1	0.1492	0.0313	4.7601	5.3338E-06
propalug-0.5	0.0075	0.0014	5.5550	1.6345E-07