INSTITUTO SUPERIOR DE ECONOMIA E GESTÃO
Estatística II - Licenciatura em Gestão - 27 Junho 2011
Parte teórica

Nome: ____ Nº ________

1. **Perguntas de Verdadeiro/Falso** (2 valores) - Para cada afirmação, assinale se esta é Verdadeira (V) ou Falsa (F). Uma resposta certa vale 0.25 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.		
No modelo de regressão $y_t = \beta_1 + \beta_2 x_t + u_t$, $t = 1,2,\cdots,n$, em que $R^2 > 0.95$ pode-se afirmar que $\beta_2 > 0$.		
Quando, num modelo de regressão linear, se introduz uma restrição linear nos parâmetros, a soma dos quadrados dos resíduos não pode diminuir.		
Com base numa amostra casual simples de dimensão $n = 4$, propôs-se, como estimador para $\mu = E(X)$, $\hat{\mu} = 0.1X_1 + 0.4X_2 + 0.4X_3 + 0.1X_4$. Este estimador é centrado.		
A potência de um teste é a probabilidade de rejeitar a hipótese nula quando ela é falsa.		
Num teste de independência do qui-quadrado a frequência esperada de observações em cada célula não é relevante.		
Para que um estimador seja consistente ele tem de ser centrado		
Utilizando o procedimento habitual, os intervalos de confiança para a variância de uma variável normal são simétricos em torno de s'^2		

2. **Perguntas de resposta múltipla** (2 valores) - Para cada pergunta escolha **a** alternativa correcta. Uma resposta certa vale 0.5 valores e uma resposta errada penaliza em 0.17 valores.

   a. Seja $W = \{(x_1, x_2, \cdots, x_n) : F_{obs} > 9.48773\}$ a região crítica de um determinado teste de hipóteses baseado numa estatística de teste com distribuição *F* de Snedecor. Observada uma amostra casual verificou-se que $F_{obs} = 0.86$. Então

   - ☐ H_0 deve ser rejeitada
   - ☐ A região de aceitação é dada por $\overline{W} = \{(x_1, x_2, \cdots, x_n) : F_{obs} > 1/9.48773\}$
   - ☐ A dimensão do teste é dada pela área por baixo da função de densidade da F e à direita de 9.48773
   - ☐ Nenhuma das alternativas anteriores é verdadeira.

   b. O método da variável fulcral é um resultado particularmente importante

   - ☐ Na estimação por intervalos
   - ☐ Nos testes de hipóteses
   - ☐ Na estimação por pontos
   - ☐ Para qualquer das alternativas

   c. Considere o seguinte quadro ANOVA decorrente da estimação de um modelo de regressão.

ANOVA

	df	*SS*	*MS*	*F*
Regression	6	254.2	42.67	10.00
Residual	100	423.6	4.24	
Total	106	677.8		

   A "estimativa dos mínimos quadrados" para a variância da variável residual do modelo, σ^2, é:

   |___| 4.24 |___| 423.6 |___| 42.67 |___| 10.00

d. 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 $t,s = 1,2,...,n$,

- ☐ $\text{cov}(u_t, u_s \mid X) = 0$ para $t \neq s$
- ☐ $\text{cov}(x_{t2}, x_{s3}) = 0$
- ☐ $\text{cov}(u_t, u_s \mid X) = \sigma^2$ para $t \neq s$ e $\text{cov}(u_t, u_s \mid X) = 0$ para $t = s$
- ☐ Todas as afirmações anteriores são falsas

3. Perguntas de desenvolvimento (2 valores) **–** Cada resposta certa vale 1 valor.

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

b. Considere que, com base no modelo $y_t = \beta_1 + \beta_2 x_{t2} + \beta_3 x_{t3} + \beta_4 x_{t4} + \beta_5 x_{t5} + u_t$, queria testar $H_0 : \beta_2 = \beta_3 \wedge \beta_4 = 1 + \beta_5$. Explique como efectuar este teste apresentando eventuais modelos auxiliares que seja necessário estimar.

Nome:____ Nº ________

Alínea	1a	1b	1c	2	3a	3b	3c	3d	3e1	3e2	T:
Cotação	15	15	15	15	20	10	15	15	10	10	P:
Classificação											**TOTAL:**

Em todos os testes de hipóteses que fizer **formule as hipóteses, indique a estatística de teste e sua distribuição** e assuma por defeito uma dimensão $\alpha = 0.05$.

1. Para a instalação de um parque eólico numa determinada região pretende-se estudar a velocidade mínima diária do vento nessa região (em km/h), X, que se admite ser uma variável aleatória com $\mu = E(X) = 3\theta / 2$, $Var(X) = 3\theta^2 / 4$ e função densidade $f_X(x|\theta) = \dfrac{3\theta^3}{x^4}$ para $x > \theta$, em que $\theta > 0$ é um parâmetro desconhecido.

   Observada uma amostra casual de 200 dias obtiveram-se os seguintes resultados para a velocidade mínima diária do vento,

Média	Variância corrigida	Mediana	Mínimo	Máximo
5.19	51.81	3.78	3.05	64.40

a) Deduza o estimador de θ pelo método dos momentos. Calcule a respectiva estimativa e comente o resultado obtido.

b) Obtenha um intervalo de confiança, a aproximadamente 95%, para μ.

c) Efectue um teste de hipóteses que lhe permita averiguar se a função distribuição para a variável aleatória X, dada por

$$F_X(x) = \begin{cases} 0 & x < 3 \\ 1 - \dfrac{27}{x^3} & x \geq 3 \end{cases}$$

se adequa aos dados recolhidos, apresentados no quadro abaixo

Velocidade mínima diária (km/h)	Até 4	4 a 5	5 a 6	6 a 8	+ de 8
Nº dias	107	50	23	16	4

2. Pretende-se saber se as expectativas positivas da evolução da economia portuguesa, são similares para os residentes em zonas urbanas e em zonas rurais. Com base num inquérito realizado junto a amostras casuais seleccionadas em cada uma das zonas obtiveram-se os seguintes resultados: nas zonas urbanas das 300 pessoas inquiridas, 111 declararam ter expectativas positivas, nas zonas rurais das 200 pessoas interrogadas, 87 declararam ter expectativas positivas. Efectuando o teste de hipóteses adequado o que pode concluir.

3. Para estudar os determinantes da capacidade pulmonar vital (*CPV*) nos jovens foi estimado o seguinte modelo de regressão linear múltipla:

$$LnCPV_t = \beta_1 + \beta_2\, ID_t + \beta_3\, LnALT_t + \beta_4\, MASC_t + \beta_5\, FUM_t + U_t$$

em que, para o $t-ésimo$ elemento da amostra, $LnCPV$ representa o logaritmo natural da capacidade pulmonar vital (volume máximo de ar expirado após uma inspiração forçada, em centímetros cúbicos), ID a sua idade, $LnALT$ o logaritmo natural da sua altura (em centímetros), $MASC$ variável que assume o valor 1 se for rapaz e 0 se for rapariga e FUM assume o valor 1 se for fumador e 0 no caso contrário.

Em todas as regressões apresentadas o regressando é $LnCPV$.

Supondo satisfeitas as hipóteses do modelo de regressão linear e considerando os dados relativos a uma amostra de 643 jovens entre os 5 e os 19 anos de idade obtiveram-se os seguintes resultados:

Regression Statistics	
Multiple R	0.89392626
R Square	0.79910417
Adjusted R Square	0.79784463
Standard Error	0.14435868
Observations	643

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	4	52.8857872	13.2214468	634.44379146	1.1372E-220
Residual	638	13.2955562	0.0208394		
Total	642	66.1813433			

	Coefficients	*Standard Error*	*t Stat*	*P-value*
Intercept	-5.21783509	0.48406579	-10.7791857	5.28616E-25
ID	0.02386441	0.00334117	7.14252829	2.50738E-12
LnALT	2.53678577	0.10109562	25.09293508	3.39653E-97
MASC	0.03940252	0.01166033	3.37919515	0.00077127
FUM	-0.04064906	0.02080769	-1.95355987	0.05118994

$$\hat{Cov}(b|X) = \begin{bmatrix} 0.2343197 & 0.0011792 & -0.0488847 & 0.0010223 & -0.0003880 \\ 0.0011792 & 0.0000112 & -0.0002560 & 0.0000042 & -0.0000213 \\ -0.0488847 & -0.0002560 & 0.0102203 & -0.0002254 & 0.0001086 \\ 0.0010223 & 0.0000042 & -0.0002254 & 0.0001360 & 0.0000200 \\ -0.0003880 & -0.0000213 & 0.0001086 & 0.0000200 & 0.0004330 \end{bmatrix}$$

a) Poder-se-á concluir que as variáveis incluídas no modelo são no seu conjunto estatisticamente significativas? Interprete o valor das estimativas obtidas para β_2 e β_4 .

b) Comente, justificando, a seguinte frase: “Com uma dimensão de 0.05 não existe evidência estatística para afirmar que fumar faz diminuir, em média, a capacidade pulmonar vital”.

c) Calcule o intervalo de confiança a 95% para β_3. Face ao resultado obtido diga se é aceitável concluir, com uma confiança de 95%, que um acréscimo relativo de 1% na altura de uma pessoa pode induzir, em média, um acréscimo superior a 2% na sua capacidade pulmonar vital, mantendo-se tudo o resto inalterado.

d) Efectue o teste estatístico $H_o : \beta_4 + \beta_5 = 0$ contra $H_1 : \beta_4 + \beta_5 \neq 0$. Interprete o seu significado.

e) Com o objectivo de prever a capacidade pulmonar vital das raparigas não fumadoras, com 15 anos de idade e uma altura de 160 centímetros estimou-se o modelo:

$$LnCPV = \beta_1 + \beta_2\, ID0 + \beta_3\, LnALT0 + \beta_4\, MASC0 + \beta_5\, FUM0 + U$$

e1) Defina $ID0$, $LnALT0$, $MASC0$ e $FUM0$ em termos dos regressores do modelo inicial.

e2) Calcule o intervalo de previsão a 95% para a capacidade pulmonar vital média dessas raparigas.

Regression Statistics	
Multiple R	0.89178954
R Square	0.79528858
Adjusted R Square	0.79432749
Standard Error	0.14560906
Observations	643

ANOVA

	Df	*SS*	*MS*	*F*	*Significance F*
Regression	3	52.63326627	17.54442209	827.4890739	1.464E-219
Residual	639	13.54807703	0.021201998		
Total	642	66.1813433			

	Coefficients	*Standard Error*	*t Stat*	*P-value*
Intercept	-5.431313204	0.484324348	-11.2142064	9.1024E-27
ID	0.019993339	0.003177967	6.291235966	5.84545E-10
LnALT	2.587347192	0.100913307	25.6393063	3.062E-100
MASC+FUM	0.021855006	0.010605458	2.060731839	0.039732797

Regression Statistics	
Multiple R	0.893926264
R Square	0.799104166
Adjusted R Square	0.797844631
Standard Error	0.144358684
Observations	643

ANOVA

	Df	*SS*	*MS*	*F*	*Significance F*
Regression	4	52.88578716	13.22144679	634.4437915	1.1372E-220
Residual	638	13.29555615	0.02083943		
Total	642	66.1813433			

	Coefficients	*Standard Error*	*t Stat*	*P-value*
Intercept	8.014759819	0.016986610	471.8280816	0
ID0	0.023864414	0.003341172	7.142528291	2.50738E-12
LnALT0	2.536785768	0.101095618	25.09293508	3.39653E-97
MASC0	0.039402518	0.011660326	3.379195152	0.000771273
FUM0	-0.040649059	0.020807685	-1.953559871	0.051189940