Estatística II – Licenciatura em Gestão
ER - 28 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 (várias situações)	X	
No modelo de regressão linear $E(y_t \mid x_t) = \beta_1 + \beta_2 x_t$, $t = 1, 2, \cdots, n$ a elasticidade de y em ordem a x é constante		X
A eficiência, tal como foi vista no curso, é uma propriedade que só se aplica a estimadores centrados	X	
Num modelo de regressão linear estimado pelos mínimos quadrados pode-se garantir que o estimador de qualquer dos parâmetros β_j é uma combinação linear dos valores observados para o regressando (y_t).	X	
A potência de um teste é igual ao seu valor-p		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. A desigualdade de Fréchet-Cramer-Rao (FCR) é um resultado teórico importante no estudo
      - ☐ Do modelo de regressão linear
      - ☐ Do enviesamento de um estimador
      - ☐ Dos testes de hipóteses
      - X Da eficiência de um estimador

   b. Seja 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. Para testar $H_0 : \beta_2 + 3\beta_3 = 0$ contra $H_1 : \beta_2 + 3\beta_3 \neq 0$ utilizando o procedimento apresentado no curso, qual das regressões auxiliares se deve utilizar?
      - X $y_t = \beta_1 + \beta_3 (x_{t3} - 3x_{t2}) + \beta_4 x_{t4} + u_t$
      - ☐ $y_t = \beta_1 + \beta_3 (x_{t3} + 3x_{t2}) + \beta_4 x_{t4} + u_t$
      - ☐ $y_t = \beta_1 + \beta_3 (x_{t3} - 3x_{t2} + x_{t4}) + u_t$
      - ☐ $y_t = \beta_1 + \beta_3 (x_{t3} - 3x_{t2} - x_{t4}) + u_t$

   c. Se um teste de hipótese simples contra hipótese simples tem potência 0.67 isto significa que
      - X Quando H_1 é verdadeira, a probabilidade do teste rejeitar H_0 é 0.67
      - ☐ Quando H_0 é verdadeira, a probabilidade do teste rejeitar H_1 é 0.67
      - ☐ Quando H_1 é verdadeira, a probabilidade do teste não rejeitar H_0 é 0.67
      - ☐ Quando H_0 é verdadeira, a probabilidade do teste não rejeitar H_1 é 0.67

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

a. Defina o conceito de estimador centrado e refira a sua utilidade

Definição 7.2 do livro – Um estimador T para o parâmetro θ diz-se centrado se e só se $E(T)=\theta$, $\forall\theta\in\Theta$ em que Θ representa o espaço-parâmetro, isto é, o conjunto dos valores admissíveis para o parâmetro.

A utilidade de um estimador ser centrado destina-se a eliminar estimadores "viciados" isto é que sub-estimem ou que sobre-estimem o parâmetro em análise.

b. Considere o MRL $y_t=\beta_1+\beta_2 x_{t2}+\beta_3 \ln x_{t3}+u_t$ a verificar as hipóteses habituais. Apresente a forma da regressão auxiliar que irá utilizar para prever o valor esperado condicionado de y quando $x_2=3$ e $x_3=12$. Refira como obteria, a partir da regressão auxiliar, uma previsão pontual para este valor esperado.

Seja então $\theta=E(y\,|\,x_2=3,x_3=12)$.
A regressão auxiliar será dada por $y_t=\theta+\beta_2(x_{t2}-3)+\beta_3(\ln x_{t3}-\ln 12)+u_t$.
Estimada a regressão auxiliar, a previsão pontual será dada por $\hat{\theta}$, estimativa MQ de θ.

Instituto Superior de Economia e Gestão
UNIVERSIDADE TÉCNICA DE LISBOA

Estatística II – Licenciatura em Gestão
ER - 28 Junho 2012

Parte Prática

Nome: ____ **Nº**________

Espaço reservado para classificações				
1. (15)	3. (15)	5a. (15)	5d. (15)	**T:**
2. (15)	4. (20)	5b. (15)	5e. (15)	**P:** _____
		5c. (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 página em branco no fim do enunciado

1. Seja X uma variável aleatória com distribuição gama de parâmetros $\alpha = 4$ e λ desconhecido, isto é, $f(x \mid \lambda) = \dfrac{\lambda^4 e^{-\lambda x} x^3}{6}$, $x > 0$. Sabendo que se observou uma amostra casual de dimensão $n = 94$ que originou $\bar{x}$ = 100, qual a estimativa de máxima verosimilhança para λ?

$$L(\lambda \mid x_1, x_2, \cdots, x_n) = \prod_{i=1}^{n} f(x_i \mid \theta)$$

$$\ell(\lambda) = \ln L(\lambda \mid x_1, x_2, \cdots, x_n) = \sum_{i=1}^{n} \ln f(x_i \mid \theta) = \sum_{i=1}^{n} \ln\left(\frac{1}{6}\lambda^4 e^{-\lambda x_i} x_i^3\right)$$

$$= \sum_{i=1}^{n}\left(-\ln 6 + 4\ln\lambda - \lambda x_i + 3\ln x_i\right)$$

$$\ell'(\lambda) = \sum_{i=1}^{n}\left(\frac{4}{\lambda} - x_i\right) = \frac{4n}{\lambda} - \sum_{i=1}^{n} x_i$$

Logo $\ell'(\lambda) = 0 \Leftrightarrow \dfrac{4n}{\lambda} = \sum_{i=1}^{n} x_i \Leftrightarrow \lambda = \dfrac{4n}{\sum_{i=1}^{n} x_i} = \dfrac{4}{\bar{x}}$

Verificado que $\ell''(\theta) = -\dfrac{4n}{\lambda^2} < 0$, tem-se para estimativa de máxima verosimilhança $\hat{\lambda} = \dfrac{4}{100} = 0.04$

2. Um analista pretende comparar os resultados líquidos das PMEs do sector das Telecomunicações com os das PMEs do sector das Tecnologias de Informação. Para esse fim, seleccionou uma amostra casual com 16 empresas do sector das Telecomunicações e outra amostra casual com 16 empresas do sector das Tecnologias de Informação, e calculou a média amostral e variância corrigida dos resultados líquidos das empresas dos dois sectores. Foram obtidos os seguintes resultados (em milhões de euros):

Telecomunicações:	$\bar{x}_1$ = 0.215	s'^2_1 = 0.024
Tecnologia de Informação:	$\bar{x}_2$ = 0.185	s'^2_2 = 0.018

Admita que os resultados líquidos das empresas dos dois sectores seguem uma distribuição normal com igual variância. Construa um intervalo de confiança a 95% para a diferença entre a média dos resultados líquidos dos dois sectores. O que pode o analista concluir?

Universos normais – Intervalo de confiança para a diferença de médias, $\mu_1 - \mu_2$, quando as variâncias, embora desconhecidas, são iguais.

Variável fulcral:

$$\frac{\bar{X}_1 - \bar{X}_2 - (\mu_1 - \mu_2)}{\sqrt{\frac{1}{m}+\frac{1}{n}}\sqrt{\frac{(m-1)S_1'^2 + (n-1)S_2'^2}{m+n-2}}} \sim t(m+n-2)$$

O intervalo de confiança vem dado por $(\bar{x}_1 - \bar{x}_2) \pm t_{\alpha/2}\, s^*$ onde $s^* = \sqrt{\frac{1}{m}+\frac{1}{n}}\sqrt{\frac{(m-1)s_1'^2 + (n-1)s_2'^2}{m+n-2}}$

Ora

$m = n = 16$; $\bar{x}_1 - \bar{x}_2 = 0.215 - 0.185 = 0.03$; $s^* = \sqrt{\frac{1}{16}+\frac{1}{16}} \times \sqrt{\frac{15 \times 0.024 + 15 \times 0.018}{30}} = 0.05123$

$t_{\alpha/2} = 2.042$ *t-student* com 30 graus de liberdade

Intervalo de confiança:

(0.03 – 2.042 * 0.05123; 0.03 + 2.042 * 0.05123) = (-0.0746; 0.1346)

O intervalo de confiança contém o valor zero. Logo, não se pode dizer que haja diferença entre os resultados líquidos médios dos dois sectores, com um nível de confiança de 95%.

3. Um psicometrista desenvolveu uma nova escala para avaliação da capacidade cognitiva de crianças de 6 anos. A escala será considerada adequada se a variância das pontuações que ela gera for de 225 para a população. Para este fim, avaliou uma amostra casual de 25 crianças de 6 anos e calculou a variância amostral corrigida, s'^2 = 258. As pontuações de testes psicométricos seguem uma distribuição normal. Com base num teste de dimensão 0.05, o que pode o psicometrista concluir?

Vai-se testar $H_0 : \sigma^2 = 225$ contra $H_1 : \sigma^2 \neq 225$

Estatística de teste:

$$Q = \frac{(n-1)S'^2}{\sigma_0^2} \sim \chi^2(n-1) \quad \text{com } n = 25$$

$$q_{obs} = \frac{(25-1)\times 258}{225} = 27.52$$

$q_{1-\alpha/2} = 12.401$; $q_{\alpha/2} = 39.364$ Qui-quadrado com 24 graus de liberdade

Região de rejeição: $W_q = \{q : q < 12.401 \vee q > 39.364\}$

Como $q_{obs} \notin W_q$, não se rejeita H_0 para uma dimensão de teste de 0.05, isto é, não se rejeita que a escala construída seja adequada.

Alternativamente poder-se-ia recorrer ao valor-p

4. A taxa de desintegração radioactiva de núcleos pesados é bem modelizada por um processo de Poisson. O número de desintegrações radioactivas de um quilograma de Urânio-235 foi medido em 100 períodos com duração de 1 segundo. As frequências observadas foram as seguintes:

Número de desintegrações:	0	1	2	3
Número de períodos:	63	27	8	2

Teste a hipótese da taxa de desintegração radioactiva de um quilograma de Urânio-235 seguir um processo de Poisson com média de 0.5 desintegrações por segundo, isto é, teste se é estatisticamente aceitável considerar que a amostra observada provém de uma distribuição de Poisson de parâmetro 0.5.

Seja X o número de desintegrações por segundo de um quilograma de Urânio-235. Vamos proceder a um teste do qui-quadrado à bondade do ajustamento

$H_0 : X \sim \text{Po}(0.5) \rightarrow H_0' : p_j = p_{0j}, j = 1, 2, \cdots, m$ contra $H_1' : p_j \neq p_{0j}$, para algum j.

Estatística de teste: $$Q = \sum_{j=1}^{m} \frac{\left(N_j - np_{0j}\right)^2}{np_{0j}} \overset{a}{\sim} \chi^2(m-1)$$

x	p_x	$n\times p_x$	n_j
0	0.6065	60.65	63
1	0.3033	30.33	27
2	0.0758	7.58	8
3 ou mais	0.0144	1.44	2

x	$n\times p_x$	n_j
0	60.65	63
1	30.33	27
≥2	9.02	10

Houve necessidade de reagrupar as 2 últimas classes já que a frequência esperada era < 5.

$q_{obs} = 0.563$; $m = 3$; $\text{valor}-\text{p} = \text{P}\left(Q > q_{\text{obs}} \mid \text{H}_0\right) = 0.755$

Não se rejeita a hipótese nula de um quilograma de Urânio-235 seguir um processo de Poisson com média (intensidade) de 0.5 desintegrações por segundo.

5. Com o objectivo de estudar o preço de venda de novos automóveis, construiu-se o seguinte modelo:

$$lpreco_t = \beta_1 + \beta_2 lconsmed_t + \beta_3 lco2_t + \beta_4 pot_t + \beta_5 velmax_t + \beta_6 acel_t + u_t.$$

Onde,

- $lpreco$ – logaritmo do preço (euros) do automóvel;
- $lconsmed$ – logaritmo do consumo médio de combustível;
- $lco2$ – logaritmo da emissão de dióxido de carbono;
- pot – potência do automóvel (em cavalos);
- $velmax$ – velocidade máxima do automóvel (em km/h);
- $acel$ – tempo de aceleração dos 0 aos 100 km/h (em segundos);

Recolhida uma amostra composta por 292 automóveis, estimou-se o modelo recorrendo ao EXCEL. Os resultados encontram-se no **Modelo 1**, em anexo.

a) Interprete as estimativas dos coeficientes dos regressores $lconsmed$ e $velmax$.

$b_2 = -2.3842$ Tudo o resto igual um aumento de 1% no consumo médio de combustível originará aproximadamente uma diminuição média de 2.3842% no preço do veículo.

$b_5 = 0.0004$ Tudo o resto igual um aumento de uma unidade (mais 1 km/h de velocidade máxima) originará aproximadamente um aumento médio de 0.04% no preço do veículo.

b) Com base no valor-p, analise a significância estatística dos regressores $lconsmed$ e $velmax$.

O teste que se irá fazer para cada um dos coeficientes ($j = 2$ ou 5) é $H_0 : \beta_j = 0$ contra $H_1 : \beta_j \neq 0$ cuja estatística de teste, sob H_0, tem distribuição $\frac{b_j}{s_{b_j}} \sim t(286)$.

Para o teste de β_2 tem-se valor-p=0.0000, logo rejeita-se H_0 e conclui-se que o regressor $lconsmed$ tem significância estatística.

Para o teste de β_5 tem-se valor-p=0.6922, logo não se rejeita H_0 e conclui-se que o regressor $velmax$ não tem significância estatística.

c) Apresente um intervalo de confiança a 90% para a elasticidade do preço em relação as emissões de dióxido de carbono.

Variável fulcral $\frac{b_3 - \beta_3}{s_{b_3}} \sim t(286)$

O intervalo é dado por $b_3 \pm t_{\alpha/2} \times s_{b_3}$ com $t_{\alpha/2} \approx 1.645$ logo tem-se (2.930, 3.659)

d) Construa um intervalo de previsão a 90% para o preço médio de um automóvel com as seguintes características: $consmed = 5$, $co2 = 100$, $pot = 300$, $velmax = 270$ e $acel = 4$. Admita que $s^2 c(X^T X)^{-1} c^T = 0.005866$.

Seja $\theta = E(lpreco \mid lconsmed = \ln(5), lco2 = \ln(100), pot = 300, vel\max = 270, acel = 4)$

Intervalo de previsão para θ : (10.2317; 10.4837) já que

$$\frac{\hat{\theta}-\theta}{\sqrt{s^2\, c(X^T X)^{-1} c}} \sim t_{(286)} \text{ logo } t_{\alpha/2} = 1.645$$

$$\hat{\theta} = -1.5491 - 2.3842 \times \ln(5) + 3.2943 \times \ln(100) + 0.0024 \times 300 + 0.0004 \times 270 - 0.0637 \times 4$$
$$= 10.3577$$

$$\sqrt{s^2\, c(X^T X)^{-1} c} = \sqrt{0.005866} = 0.0766$$

Intervalo de previsão para $E(preco)$: $\left(e^{10.2317}; e^{10.4837}\right)$ isto é (27769.7; 35727.7)

f) Posteriormente, definiu-se a variável binária d que assume o valor 1 se o automóvel é todo-o-terreno e assume o valor 0 se o automóvel é ligeiro. Suponha agora a seguinte regressão, em que se adiciona ao modelo inicial interacções entre a variável binária e todos os regressores:

$$lpreco_t = \beta_1 + \delta_1 d_t + \beta_2 lconsmed_t + \delta_2 d_t lconsmed_t + \beta_3 lco2_t + \delta_3 d_t lco2_t + \beta_4 pot_t +$$
$$+\delta_4 d_t pot_t + \beta_5 velmax_t + \delta_5 d_t velmax_t + \beta_6 acel_t + \delta_6 d_t acel_t + u_t.$$

Do que se está a desconfiar ao formular este modelo?

Os resultados da estimação deste modelo encontram-se no **Modelo 2** em anexo. Efectuando o teste de hipóteses adequado ($\alpha = 0.05$) diga se a suspeita se confirma.

A regressão apresentada nesta alínea destina a investigar se o modelo explicativo do preço dos veículos todo-o-terreno é o mesmo do que para os restantes ligeiros (existência ou não de permanência de estrutura)

No teste de hipótese que se vai fazer, H_0 corresponde à permanência de estrutura enquanto H_1 corresponde a uma quebra de estrutura.

$H_0 : \delta_1 = \delta_2 = \cdots = \delta_6 = 0$ contra $H_1 : \exists \delta_j \neq 0, j = 1, 2, \cdots, 6$

Estatística de teste:

$F = \dfrac{(VR_0 - VR_1)/6}{VR_1/280} \sim F(6;280)$ sendo VR_0 a variação residual do modelo com restrições (modelo 1 do anexo) e VR_1 a variação residual do modelo sem restrições (modelo 2 do anexo).

$$F_{obs} = \frac{(13.3562 - 11.3776)/6}{11.3776/280} = 8.115$$

Como $F_{0.05} = 2.10$, rejeita-se H_0 e conclui-se que o modelo explicativo do preço dos veículos todo-o-terreno não é o mesmo do que para os restantes ligeiros (existe quebra de estrutura).

Modelo 1

Regression Statistics	
Multiple R	0.9435
R Square	0.8901
Adjusted R Square	0.8882
Standard Error	0.2161
Observations	292

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	5	108.1818	21.6364	463.3052	0.0000
Residual	286	13.3562	0.0467		
Total	291	121.5380			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	-1.5491	0.8877	-1.7451	0.0820	-3.2963	0.1981
lconsmed	-2.3842	0.1965	-12.1337	0.0000	-2.7709	-1.9974
lco2	3.2943	0.2214	14.8767	0.0000	2.8585	3.7302
pot	0.0024	0.0003	7.1259	0.0000	0.0017	0.0030
velmax	0.0004	0.0011	0.3962	0.6922	-0.0017	0.0026
acel	-0.0637	0.0133	-4.8003	0.0000	-0.0898	-0.0376

Modelo 2

Regression Statistics	
Multiple R	0.9520
R Square	0.9064
Adjusted R Square	0.9027
Standard Error	0.2016
Observations	292

ANOVA

	df	*SS*	*MS*	*F*
Regression	11	110.1604	10.0146	246.4557
Residual	280	11.3776	0.0406	
Total	291	121.5380		

	Coefficients	*Standard Error*	*t Stat*	*P-value*
Intercept	4.0732	1.3933	2.9235	0.0037
d	-5.9624	1.9085	-3.1241	0.0020
lconsmed	-1.4890	0.2947	-5.0524	0.0000
d*lconsmed	-0.6719	0.4088	-1.6435	0.1014
lco2	1.6316	0.3893	4.1910	0.0000
d*lco2	1.5520	0.5032	3.0844	0.0022
pot	0.0029	0.0005	5.4348	0.0000
d*pot	-0.0010	0.0007	-1.3708	0.1715
velmax	0.0041	0.0015	2.6997	0.0074
d*velmax	-0.0016	0.0023	-0.7046	0.4817
acel	-0.0527	0.0162	-3.2532	0.0013
d*acel	0.0079	0.0262	0.3007	0.7639