INSTITUTO SUPERIOR DE ECONOMIA E GESTÃO

UNIVERSIDADE DE LISBOA

ESTATÍSTICA II – LICENCIATURA EM GESTÃO

Exame de Época Normal – 13 de Janeiro de 2014

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**
Sejam θ e φ dois parâmetros desconhecidos, tais que $\theta = g(\varphi)$, sendo g uma função. Se $\hat{\varphi}$ é um estimador de máxima verosimilhança para φ, então $\hat{\theta} = g(\hat{\varphi})$ é necessariamente um estimador de máxima verosimilhança para θ.		X
Se num teste estatístico obtiver um valor-p de 0.03, deve rejeitar-se H_0 num teste com dimensão de 5%, mas não deve rejeitar-se H_0 num teste com dimensão de 1%.	X	
No teste de independência entre dois atributos, A e B, de uma população, o número observado de elementos em cada célula (A_i, B_j) deverá ser de pelo menos 5.		X
No modelo $y_t = \beta_1 + \beta_2 x_{t2} + \beta_3 x_{t3} + u_t$, quanto maior for o grau de multicolinearidade entre x_{t2} e x_{t3}, menor será a variância dos estimadores de MQ de β_2 e β_3, tudo o resto igual.		X
No modelo de regressão linear não é usual testar a significância individual do termo independente.	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. Num teste de hipóteses é cometido um erro de segunda espécie quando:

☐ Rejeita-se a hipótese alternativa quando esta é verdadeira;

☐ Rejeita-se a hipótese nula quando esta é verdadeira;

☐ Não se rejeita a hipótese alternativa quando esta é falsa;

X Não se rejeita a hipótese nula quando esta é falsa.

b. Considere o modelo de regressão linear $\hat{y}_t = 800 + 50d_t + 30x_t + 20d_t x_t$, onde y_t é o salário mensal (em €) do trabalhador, d_t é uma variável binária que assume o valor 1 se o trabalhador for do sexo masculino e o valor 0 se for do sexo feminino, e x_t é uma variável quantitativa que representa a experiência profissional (em anos). Podemos dizer que, por cada ano adicional de experiência profissional, a diferença entre o salário médio estimado dos homens e o salário médio estimado das mulheres aumenta:

X 20 euros

☐ 30 euros

☐ 50 euros

☐ 70 euros

c. Num modelo de regressão linear, a hipótese de homocedasticidade implica que:

☐ A média da variável residual seja zero;

☐ A variável residual tem distribuição normal;

X A variância da variável residual é constante;

☐ A correlação entre as variáveis residuais das diferentes observações é zero.

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

a) Seja (X_1, X_2, X_3) uma amostra aleatória extraída de uma população com média μ e variância σ^2. Considere os seguintes estimadores pontuais para μ:

$$T_1 = \frac{1}{8}X_1 + \frac{3}{4}X_2 + \frac{1}{8}X_3, \qquad T_2 = \frac{1}{4}X_1 + \frac{1}{2}X_2 + \frac{1}{4}X_3$$

Mostre que ambos os estimadores são centrados e identifique, justificando, o mais eficiente.

$$E(T_1) = \frac{1}{8}E(X_1) + \frac{3}{4}E(X_2) + \frac{1}{8}E(X_3) = \frac{1}{8}\mu + \frac{3}{4}\mu + \frac{1}{8}\mu = \mu$$

$$E(T_2) = \frac{1}{4}E(X_1) + \frac{1}{2}E(X_2) + \frac{1}{4}E(X_3) = \frac{1}{4}\mu + \frac{1}{2}\mu + \frac{1}{4}\mu = \mu$$

Ambos os estimadores são centrados.

$$Var(T_1) = \frac{1}{64}Var(X_1) + \frac{9}{16}Var(X_2) + \frac{1}{64}Var(X_3) = \frac{1}{64}\sigma^2 + \frac{9}{16}\sigma^2 + \frac{1}{64}\sigma^2 = \frac{38}{64}\sigma^2 = 0.594\sigma^2$$

$$Var(T_2) = \frac{1}{16}Var(X_1) + \frac{1}{4}Var(X_2) + \frac{1}{16}Var(X_3) = \frac{1}{16}\sigma^2 + \frac{1}{4}\sigma^2 + \frac{1}{16}\sigma^2 = \frac{6}{16}\sigma^2 = 0.375\sigma^2$$

O estimador T_2 é mais eficiente do que T_1.

b) Suponha que, após a estimação do modelo de regressão linear $y_t = \beta_1 + \beta_2 x_{t2} + \cdots + \beta_k x_{tk} + u_t$ obtém a seguinte informação: $VR = \sum_{t=1}^{n} \hat{u}_t^2 = 100$ e $VE = \sum_{t=1}^{n} (\hat{y}_t - \bar{y})^2 = 200$. Qual é o valor do coeficiente de correlação empírico entre a variável explicada, y, e os valores previstos pelo modelo, $\hat{y}$?

$$VT = VE + VR = 300$$

$$R^2 = \frac{VE}{VT} = \frac{2}{3}$$

No caso do MRLC com termo independente:

$$r_{y\hat{y}}^2 = R^2 = \frac{2}{3} \Leftrightarrow r_{y\hat{y}} = \sqrt{\frac{2}{3}} = 0.8165$$

INSTITUTO SUPERIOR DE ECONOMIA E GESTÃO

UNIVERSIDADE DE LISBOA

ESTATÍSTICA II – LICENCIATURA EM GESTÃO

Exame de Época Normal – 13 de Janeiro de 2014

PARTE PRÁTICA

Nome: ____ **Nº**___________

Espaço reservado para classificações				
1. (20)	2. (10)	3. (15)	4. (20)	**T:**
5a. (15)	5b. (20)	5c. (20)	5d. (20)	**P:**

ATENÇÃO! 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.

1. De uma população com função densidade de probabilidade $f(x) = \theta x^{\theta-1}, 0 < x < 1, \theta > 0$, foi obtida uma amostra aleatória com 4 observações: $x_1 = 0.05,\ x_2 = 0.15, x_3 = 0.4,\ x_4 = 0.6$. Encontre uma estimativa pelo método dos momentos para o parâmetro θ.

Primeiro momento na população:

$$\mu = E(X) = \int_0^1 xf(x)dx = \int_0^1 \theta x^\theta dx = \left|\frac{\theta x^{\theta+1}}{\theta+1}\right|_0^1 = \frac{\theta}{\theta+1}$$

Primeiro momento na amostra:

$$\bar{x} = \frac{0.05 + 0.15 + 0.4 + 0.6}{4} = 0.3$$

Assim:

$$\frac{\tilde{\theta}}{\tilde{\theta}+1} = 0.3 \ \Leftrightarrow\ \tilde{\theta} = \frac{3}{7} = 0.4286$$

2. Num inquérito a 200 alunos do ISEG que frequentaram a disciplina de Estatística II, 65 admitiram ter adormecido pelo menos uma vez nas aulas teóricas dessa disciplina. Construa um intervalo de confiança a 95% para a proporção de alunos do ISEG que já adormeceram nas aulas teóricas de Estatística II.

Variável fulcral:

$$\frac{\bar{X}-\theta}{\sqrt{\frac{\bar{X}(1-\bar{X})}{n}}} \overset{a}{\sim} N(0,1)$$

Intervalo de confiança: $\left(\bar{x}-z_{\alpha/2}\sqrt{\frac{\bar{x}(1-\bar{x})}{n}}, \bar{x}-z_{\alpha/2}\sqrt{\frac{\bar{x}(1-\bar{x})}{n}}\right)$

Assim: $\bar{x}=\frac{65}{200}=0.325$; $z_{0.025}=1.96$; $IC=(0.2589, 0.3911)$

3. Suponha que, para uma população com média μ e variância $\sigma^2=100$, pretende testar, com dimensão de teste de 0.05, as seguintes hipóteses nula e alternativa: $H_0\!: \mu=5$ contra $H_1\!: \mu\neq 5$. Se tiver uma amostra aleatória com 400 observações, quais são os valores da média amostral que implicam uma rejeição da hipótese nula?

Estatística de teste para a média de uma população (grandes amostras):

$$Z=\frac{\bar{X}-\mu_0}{\sigma/\sqrt{n}} \overset{a}{\sim} N(0,1)$$

Região de rejeição:

$$W_Z=\{z : z<-z_{\alpha/2} \vee z>z_{\alpha/2}\} \Leftrightarrow W_{\bar{X}}=\left\{\bar{x} : \bar{x}<\mu_0-z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \vee \bar{x}>\mu_0+z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right\}$$

Logo:

$$W_{\bar{X}}=\left\{\bar{x} : \bar{x}<5-1.96\frac{10}{\sqrt{400}} \vee \bar{x}>5+1.96\frac{10}{\sqrt{400}}\right\}=\{\bar{x} : \bar{x}<\ 4.02 \vee \bar{x}>5.98\}$$

4. Uma análise histórica à duração dos filmes de longa metragem indica que 9% têm uma duração inferior a 90 minutos, 34% têm uma duração entre 90 e 120 minutos, 36% têm uma duração entre 120 e 150 minutos, 18% têm uma duração entre 150 e 180 minutos, e 3% têm uma duração superior a 180 minutos. Numa amostra aleatória de 100 filmes de longa metragem produzidos em 2013, 6 têm uma duração inferior a 90 minutos, 51 têm uma duração entre 90 e 120 minutos, 21 têm uma duração entre 120 e 150 minutos, 15 têm uma duração entre 150 e 180 minutos, e 7 têm uma duração superior a 180 minutos. Teste se a duração dos filmes produzidos em 2013 segue o padrão histórico.

$H_0 : p_1 = 0.09 \wedge p_2 = 0.34 \wedge p_3 = 0.36 \wedge p_4 = 0.18 \wedge p_5 = 0.03$

$H_1 : p_1 \neq 0.09 \vee p_2 \neq 0.34 \vee p_3 \neq 0.36 \vee p_4 \neq 0.18 \vee p_5 \neq 0.03$

duração	**%**	**E**	**A**	**E**	**A**	**$(A-E)^2/E$**
<90	0.09	9	6	9	6	1.00
90-120	0.34	34	51	34	51	8.50
120-150	0.36	36	21	36	21	6.25
150-180	0.17	18	15	21	22	0.05
>180	0.04	3	7			
	1.00	100	100	100	100	15.80

O valor crítico é: $\chi^2_{3,0.05}$ = 7.815

Logo, rejeita-se a hipótese nula. Há evidência de que a duração dos filmes produzidos em 2013 não segue o padrão histórico.

5. Uma organização para a promoção da eficiência energética desenvolveu um modelo econométrico que pretende explicar o consumo de um automóvel em litros/100 km ($cons$) em função do número de cilindros do motor ($cili$), da potência do automóvel em cavalos (cv), do peso do automóvel em kg ($peso$) e da aceleração dos 0 aos 100 km/h em segundos ($acel$). Com base numa amostra casual de 392 modelos foi estimado o **Modelo 1**, que se encontra no Anexo.

a) Interprete a estimativa para o coeficiente de regressão da variável cv e teste a sua significância.

$b_3 = 0.0440$ Ceteris paribus, por cada cavalo adicional na potência do automóvel, o consumo aumenta, em média, 0.044 litros/100km.

$H_0: \beta_3 = 0$ contra $H_1: \beta_3 \neq 0$

Estatística de teste:

$$T = \frac{b_3 - \beta_3}{s_{b_3}} \sim t(n-k)$$

Valor-p = 0.0000, há evidência de que β_3 é significativo para uma dimensão de teste de 0.05.

b) Sabendo-se que $\widehat{Cov}(b_3, b_5) = 0.0002$, teste a restrição $\beta_5 = 3\beta_3$.

$$H_0 : \beta_5 = 3\beta_3 \quad H_1 : \beta_5 \neq 3\beta_3$$

$$t = \frac{\hat{\delta} - \delta}{s_{\hat{\delta}}} \sim t((n-k)$$

$$\hat{\delta} = b_5 - 3b_3 = 0.1262 - 3 \times 0.0440 = -0.0058, \qquad \delta = \beta_5 - 3\beta_3 = 0$$

$$s_{\hat{\delta}} = \sqrt{9 \times 0.0063^2 + 0.0490^2 - 6 \times 0.0002} = 0.0385 \quad t_{obs} = \frac{-0.0058}{0.0385} = -0.1518$$

Uma vez que $W_T = \{t : t < 1.96 \vee t > 1.96\}$, não se rejeita a restrição proposta.

c) Um novo modelo de automóvel tem um motor com 4 cilindros, uma potência de 100 cavalos, um peso de 1200 kg e uma aceleração de 11 segundos dos 0 aos 100 km. Apresente um intervalo de previsão a 95% para o consumo deste modelo de automóvel.

A previsão pontual do consumo é dada pela estimativa do coeficiente independente: $\hat{y}_0 = 9.3508$

Por outro lado, o erro padrão da previsão em média é dado pelo erro padrão do termo independente: $s_{\hat{\theta}} = 0.2702$

Uma vez que $s^2 = 2.7711$ temos

$$s_d = \sqrt{s^2 + s_{\hat{\theta}}^2} = \sqrt{2.7711 + 0.2702^2} = 1.6864$$

Logo:

$$\left(\hat{y}_0 - t_{387,0.025} s_d, \hat{y}_0 + t_{387,0.025} s_d\right) = (9.3508 - 1.96 \times 1.6864, 9.3508 + 1.96 \times 1.6864) = (6.0455, 12.6561)$$

d) Posteriormente, estimaram-se duas regressões: uma com os dados dos 307 modelos automóvel com data de construção anterior ao ano 2000; outra com os 85 modelos automóvel restantes. Os resultados da estimação correspondem aos **Modelos 3**, que se encontram no Anexo. Indique o nome do teste que poderá ser realizado com esta regressão, efectue esse teste e apresente as suas conclusões.

Teste de alteração de estrutura.

$$H_0 : \ \beta_{j1} = \beta_{j2} \forall j = 1, \ldots, 5 \ \ H_1 : \ \exists j = 1, \ldots, 5 \ \ \beta_{j1} \neq \beta_{j2}$$

Estatística de teste:

$$F_{Chow} = \frac{(VR_0 - VR_1)/k}{VR_1/(n - 2k)} \sim F(k, n - 2k)$$

$$VR_0 = 1072.4143, \qquad VR_1 = 834.9775\ , \qquad n = 392, \qquad k = 5$$

$$F_{obs} = 21.7254, \qquad F_{5,382,0.05} = 2.21$$

Rejeita-se a hipótese nula. Há evidência de alteração de estrutura entre o modelo explicativo do consumo dos modelos automóvel anteriores ao ano 2000 e o modelo explicativo do consumo dos modelos automóvel restantes.

CONTINUAÇÃO DA QUESTÃO _____

ANEXO

Modelo 1: $cons_t = \beta_1 + \beta_2 cili_t + \beta_3 cv_t + \beta_4 peso_t + \beta_5 acel_t + u_t$

Regression Statistics	
Multiple R	0.9060
R Square	0.8208
Adjusted R Square	0.8189
Standard Error	1.6647
Observations	392

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	4	4910.7935	1227.6984	443.0371	0.0000
Residual	387	1072.4143	2.7711		
Total	391	5983.2078			

	Coefficients	*Standard Error*	*t Stat*	*P-value*
Intercept	-2.8621	0.9608	-2.9789	0.0031
cili	0.3256	0.1190	2.7358	0.0065
cv	0.0440	0.0063	6.9834	0.0000
peso	0.0043	0.0006	6.7183	0.0000
acel	0.1262	0.0490	2.5748	0.0104

Modelo 2: $cons_t = \beta_1 + \beta_2(cili_t - 4) + \beta_3(cv_t - 100) + \beta_4(peso_t - 1200) + \beta_5(acel_t - 11) + u_t$

Regression Statistics	
Multiple R	0.9060
R Square	0.8208
Adjusted R Square	0.8189
Standard Error	1.6647
Observations	392

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	4	4910.7935	1227.6984	443.0371	0.0000
Residual	387	1072.4143	2.7711		
Total	391	5983.2078			

	Coefficients	*Standard Error*	*t Stat*	*P-value*
Intercept	9.3508	0.2702	34.6073	0.0000
cili - 4	0.3256	0.1190	2.7358	0.0065
cv - 100	0.0440	0.0063	6.9834	0.0000
peso - 1200	0.0043	0.0006	6.7183	0.0000
acel - 11	0.1262	0.0490	2.5748	0.0104

Modelos 3:

$$cons_t = \beta_1 + \beta_2 cili_t + \beta_3 cv_t + \beta_4 peso_t + \beta_5 acel_t + u_t, t = 1, \dots, 307$$

Regression Statistics	
Multiple R	0.9105
R Square	0.8290
Adjusted R Square	0.8268
Standard Error	1.5713
Observations	307

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	4	3616.1739	904.0435	366.1496	0.0000
Residual	302	745.6546	2.4691		
Total	306	4361.8285			

	Coefficients	*Standard Error*	*t Stat*	*P-value*
Intercept	-2.7029	1.0562	-2.5591	0.0110
cili	0.2174	0.1253	1.7351	0.0837
cv	0.0430	0.0063	6.8307	0.0000
peso	0.0044	0.0007	6.7600	0.0000
acel	0.1757	0.0532	3.3053	0.0011

$$cons_t = \beta_1 + \beta_2 cili_t + \beta_3 cv_t + \beta_4 peso_t + \beta_5 acel_t + u_t, t = 308, \dots, 392$$

Regression Statistics	
Multiple R	0.7587
R Square	0.5756
Adjusted R Square	0.5544
Standard Error	1.0567
Observations	85

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	4	121.1428	30.2857	27.1247	0.0000
Residual	80	89.3228	1.1165		
Total	84	210.4656			

	Coefficients	*Standard Error*	*t Stat*	*P-value*
Intercept	2.4915	1.4871	1.6753	0.0978
cili	-0.0405	0.1885	-0.2151	0.8303
cv	0.0113	0.0148	0.7630	0.4477
peso	0.0054	0.0012	4.6537	0.0000
acel	-0.0968	0.0725	-1.3356	0.1855