Instituto Superior de Economia e Gestão

Universidade de Lisboa

Licenciaturas em Economia e em Finanças

Estatística II – Época Normal – ___/____/2014 – Duração 2 horas

Nome:____ Número:__________

Notas: Os telemóveis devem estar desligados. O não cumprimento desta norma é motivo suficiente para anulação da prova. As perguntas de escolha múltipla têm cotação de **1** valor; **respostas erradas serão penalizadas** em **0.2** valores. Não havendo informação em contrário, deverá utilizar um nível de significância de 5%. Pode usar a última página para continuar qualquer questão. <u>Formalize devidamente todas as respostas</u>.

1. [**2.0**] Num estudo sobre a relação entre o nível de *stress* e os hábitos tabágicos dos indivíduos, foi recolhida uma amostra aleatória de 300 pessoas. Depois de medido o seu nível de *stress* assim como o número médio de cigarros fumados por dia, obteve-se o seguinte quadro de valores,

Nível de Stress	***Nº médio de cigarros, por dia***		
	Nenhum	**Entre 1 e 2**	**Mais de 2**
Abaixo da média	45	60	55
Acima da média	20	55	65

Podemos concluir que o nível de *stress* é independente dos hábitos tabágicos?

2. [**2.0**] O laboratório farmacêutico "TrataTudo" desenvolveu um novo medicamento para combater os efeitos nefastos do *stress*. Por forma a testar a eficácia deste medicamento foram escolhidos aleatoriamente 300 indivíduos, que posteriormente foram divididos, aleatoriamente, em 2 grupos. Ao grupo I, com 200 pessoas, foi ministrado o novo medicamento; enquanto que ao grupo II, com as restantes pessoas, foi ministrado o medicamento mais vendido no mercado. O nível de *stress* dos indivíduos em cada grupo foi medido através de um índice, que varia entre 0 e 100 (quanto maior o índice, maior será o nível de *stress* demonstrado pelo indivíduo). Obtiveram-se os seguintes resultados das medições efetuadas:

	Média	**Desvio-padrão**
Grupo I	$\bar{x}_1 = 70.8$	$s_1 = 12.7$
Grupo II	$\bar{x}_2 = 74.2$	$s_2 = 20.4$

Será que, perante estes dados, o laboratório "TrataTudo" pode concluir que, em média, o seu medicamento é mais eficaz? Justifique recorrendo a um teste estatístico adequado.

3. [**2.0**] Considere uma variável aleatória contínua, X, com função densidade de probabilidade,

$$f\left(x|\theta\right) = \left(1+\theta\right)\left(1-x\right)^{\theta}, \text{ com } 0 < x < 1 \text{ e } \theta > -1$$

Supondo que se observou uma amostra casual de dimensão n da população X, $\left(X_1, X_2, \ldots, X_n\right)$, deduza o estimador da máxima verosimilhança para o parâmetro θ.

[**1.0**] Considerem-se dois estimadores T_1 e T_2 para um parâmetro desconhecido, de uma dada população X . Das afirmações seguintes, assinale a correta:

- [ ] Se $Var(T_1) > Var(T_2)$ então T_1 é mais eficiente que T_2
- [ ] Se T_1 e T_2 forem estimadores centrados, então $EQM(T_1) = EQM(T_2)$
- [ ] Se T_2 for um estimador consistente, então será certamente centrado
- [ ] Se T_1 for centrado e $EQM(T_1) < EQM(T_2)$ então necessariamente $Var(T_1) < Var(T_2)$
- [ ] Nenhuma das opções anteriores

4. [**1.0**] Num teste de hipóteses, conhecendo a probabilidade de tomar uma decisão correta e sendo esta igual a 90%, significa que:

- [ ] Nada se pode concluir relativamente à dimensão e potência do teste
- [ ] A dimensão do teste, α , é igual a 10%
- [ ] A potência do teste, β , é igual a 90%
- [ ] A probabilidade associada ao erro de segunda espécie é igual a 90%
- [ ] Nenhuma das opções anteriores

5. [**1.0**] Considerando a função potência, $\beta(\theta)$, de um teste de hipóteses simples contra alternativa unilateral direita, qual das seguintes afirmações está correta:

- [ ] Esta função converge mais rapidamente para 1, à medida que a dimensão do teste, α, diminui, para uma dada dimensão da amostra, n
- [ ] Esta função converge mais rapidamente para 0, à medida que a dimensão da amostra, n, aumenta, para uma dada dimensão do teste, α
- [ ] Esta função converge mais rapidamente para 1, à medida que a dimensão da amostra aumenta, para uma dada dimensão do teste, α
- [ ] A função potência, $\beta(\theta)$, não depende de α, dimensão do teste, nem de n, dimensão da amostra
- [ ] Nenhuma das opções anteriores

[**1.0**] Considere-se o teste do ajustamento:

- [ ] Se a diferença entre as frequências esperadas e observadas for suficientemente grande, tendemos a rejeitar a hipótese nula
- [ ] Os graus de liberdade da distribuição da estatística de teste dependem do número de observações da amostra
- [ ] Se o valor observado da estatística de teste for negativo, então certamente que rejeitamos a hipótese nula
- [ ] Quando o valor das frequências observadas é suficientemente pequeno, tipicamente tendemos a rejeitar a hipótese nula
- [ ] Nenhuma das opções anteriores

6. Para explicar o salário médio dos trabalhadores de uma dada empresa considerou-se o seguinte modelo:

$$sal_i = \beta_0 + \beta_1 escol_i + \beta_2 exper_i + \beta_3 \log(QI_i) + u_i, \; i = 1, 2, \ldots, n$$

onde sal_i é o salário médio, em milhares de euros, do trabalhador i, $escol_i$ é o número de anos de escolaridade, $exper_i$ é o número de anos de experiência profissional e $\log(QI_i)$ é o logaritmo do quociente de inteligência. Após a estimação do modelo acima, obtiveram-se os seguintes resultados:

Equação 1

Dependent Variable: SAL
Method: Least Squares
Sample: 1 526
Included observations: 526

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.651886	4.756793	0.137043	0.8910
ESCOL	0.646750	0.053896	11.99991	0.0000
EXPER	0.070388	0.010986	6.407273	0.0000
LOG(QI)	-0.879748	1.021677	**(i)**	[redacted]

R-squared	0.226261	Mean dependent var	5.896103
Adjusted R-squared	0.221814	S.D. dependent var	3.693086
S.E. of regression	3.257849	Akaike info criterion	5.207587
Sum squared resid	5540.290	Schwarz criterion	5.240023
Log likelihood	-1365.595	Hannan-Quinn criter.	5.220287
F-statistic	50.88210	Durbin-Watson stat	1.825171
Prob(F-statistic)	0.000000		

[**1.0**] Interprete as estimativas obtidas para os coeficientes β_2 e β_3.

a) [**2.0**] Preencha o valor **(i)** no *output* da equação 1. Que pode concluir com a informação dada por este valor?

b) **[1.0]** Com base em informação mais recente, sabe-se que a empresa em análise contratou um novo trabalhador com $escol = 18$, $exper = 2$ e $QI = 125$. Estime o salário esperado deste trabalhador.

c) [**1.0**] Pretende-se construir um intervalo, com 95% de confiança, para a média dos salários dos trabalhadores com $escol = 18$, $exper = 2$ e $QI = 125$. Considerando o *output* da equação 1, qual das seguintes afirmações é verdadeira?

- [ ] A amplitude do intervalo é igual a $2 \times 1.96 \times 0.2263$
- [ ] Não é possível calcular o intervalo de confiança, com base no *output* da equação 1
- [ ] A amplitude do intervalo é igual a $2 \times 1.96 \times 3.2578$
- [ ] A amplitude do intervalo é igual a $2 \times 1.96 \times 4.7568$
- [ ] Nenhuma das opções anteriores

d) [**1.0**] Pretende-se realizar um teste sobre as seguintes hipóteses:

$$\begin{cases} H_0 : \beta_1 = 10\beta_2 \wedge \beta_1 - \beta_3 = 0 \\ H_1 : \beta_1 \neq 10\beta_2 \vee \beta_1 - \beta_3 \neq 0 \end{cases}$$

Para o efeito, que regressão necessitaria de conhecer, para além da equação 1?

- [ ] $sal_i = \beta_0 + \beta_1 \left[escol_i + 0.1 \times exper_i + \log\left(QI_i\right) \right] + u_i, \ i = 1, 2, \ldots, n$
- [ ] $sal_i = \beta_0 + \beta_2 \left[10 \times escol_i + exper_i + \log\left(QI_i\right) \right] + u_i, \ i = 1, 2, \ldots, n$
- [ ] $sal_i = \beta_0 + \beta_1 \left[escol_i + 0.1 \times exper_i - \log\left(QI_i\right) \right] + u_i, \ i = 1, 2, \ldots, n$
- [ ] $sal_i = \beta_0 + \beta_2 \left[escol_i + exper_i - 10 \times \log\left(QI_i\right) \right] + u_i, \ i = 1, 2, \ldots, n$
- [ ] Nenhuma das opções anteriores

e) [**1.0**] Foi testada a hipótese de presença de heterocedasticidade na equação 1, obtendo-se o seguinte resultado:

F-statistic	6.963084	Prob. F(9,516)	0.0000
LM-statistic	56.96402	Prob. Chi-Square(9)	0.0000

Das seguintes afirmações, opte pela **falsa**.

- ☐ As estatísticas t e F, obtidas com o OLS habitual, não seguem as distribuições convencionais
- ☐ As estimativas OLS são centradas, uma vez que a hipótese MLR.4 não é violada
- ☐ Os graus de liberdade da estatística de teste indicam que se trata de um teste BP habitual
- ☐ O WLS é mais eficiente que o OLS
- ☐ Não existem indícios estatísticos que nos permitam rejeitar a hipótese de heterocedasticidade

f) [**1.0**] A omissão de uma variável relevante no modelo especificado inicialmente:

- ☐ Pode ser detetada através de um teste t, de significância individual, sobre os regressores da equação 1
- ☐ Põe em causa a hipótese MLR.4, se a variável omitida estiver correlacionada com os outros regressores da equação 1
- ☐ Torna inevitavelmente o estimador OLS dos coeficientes da equação 1 enviesado
- ☐ Motiva a que as variâncias dos coeficientes da equação 1 devam ser estimadas com recurso ao estimador robusto de White
- ☐ Nenhuma das opções anteriores

g) [**2.0**] Para explicar o salário médio de um trabalhador, Mincer em 1974 sugere que o (logaritmo do) salário é uma função quadrática da experiência profissional. Ainda que a variável dependente não esteja logaritmizada no modelo da equação 1, pretendemos testar a hipótese de que a sua forma funcional pode estar a ignorar incorretamente este termo. Para o efeito, recorreu-se a um teste estatístico, cujos resultados foram os seguintes:

h)

Dependent Variable: SAL
Method: Least Squares
Included observations: 526

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	2.150626	4.655852	0.461919	0.6443
ESCOL	-0.436073	0.218569	-1.995130	0.0465
EXPER	-0.055656	0.026924	-2.067149	0.0392
LOG(QI)	0.978809	1.062351	0.921362	0.3573
FITTED^2	0.148849	0.029161	5.113402	0.0000

R-squared	0.263113	Mean dependent var	5.896103
Adjusted R-squared	0.257456	S.D. dependent var	3.693086
S.E. of regression	3.182370		
Sum squared resid	5276.415		
Log likelihood	-1352.761		

Com FITTED os valores ajustados para o salário, SAL, na Equação 1.

Nomeie e formalize o teste em causa. O que pode concluir?

Continuação da questão n.º ________