



ESTATÍSTICA II – Lic. Economia e Finanças
13 de Janeiro de 2017 – Duração: 1h

TESTE I

Nome: _____ N.º: _____

<i>Espaço reservado a classificações</i>							
1.a)	1.b)	2.a)	2.b)	2.c)	3.)	4.)	
							Total

Perguntas de escolha múltipla: cada resposta certa vale 10 pontos; cada resposta errada vale -2.5 pontos; assinale a resposta escolhida com uma cruz no quadrado adequado. As cotações das restantes perguntas são indicadas no enunciado.

1. Seja θ a proporção de alunos do 2º ano que concordam com o novo método de avaliação adoptado em certas disciplinas de determinada Universidade. O Conselho Pedagógico afirma que pelo menos 90% dos estudantes estão satisfeitos com o novo método. Para testar a veracidade desta afirmação, foram seleccionados ao acaso 200 alunos do 2º ano tendo 175 concordado com o novo método.
 - a) [15] Teste, ao nível de 5%, a veracidade da afirmação.

- b) [20] No semestre seguinte, numa amostra casual de 250 alunos, 235 alunos concordaram com o novo método de avaliação. Teste, ao nível de 5% se existem diferenças significativas, entre os dois semestres, na proporção de alunos favoráveis ao novo método de avaliação.

2. Admita que a quantidade de sumo de laranja, em litros, por pacote da nova marca *Portukal* é uma variável aleatória com distribuição normal de média μ e variância σ^2 . Com o objectivo de testar a afirmação “a quantidade média de sumo da nova marca é igual a 1”, foi recolhida uma amostra aleatória de 25 pacotes, tendo-se obtido uma média de 0.98 litros e uma variância corrigida de 0.16.
- a) [15] Construa um intervalo de confiança a 95% para a quantidade média de sumo por pacote. É de admitir, ao nível de 5%, que a afirmação está correcta? Justifique a sua resposta com base no intervalo obtido.

- b) Admita que foi construído um segundo intervalo de confiança a 95% para a quantidade média de sumo por pacote com base numa amostra casual da mesma população de dimensão superior a 25. Então:

a amplitude do segundo intervalo é menor	
a amplitude do segundo intervalo é maior	
os dois intervalos têm a mesma amplitude	
nada se pode concluir porque a amplitude não depende da dimensão da amostra	

- c) A estimativa da máxima verosimilhança para a proporção de pacotes com menos de 1 litro de sumo é:

$\Phi\left(\frac{1-\mu}{\sigma/\sqrt{n}}\right)$	
$\Phi\left(\frac{1-\bar{x}}{s}\right)$	

$\Phi\left(\frac{1-\bar{x}}{s/\sqrt{n}}\right)$	
$\Phi\left(\frac{\bar{x}-\mu}{\sigma}\right)$	

3. Seja (X_1, X_2, \dots, X_n) uma amostra casual simples retirada de uma população X cuja distribuição depende de um parâmetro θ . Admita que T_1 e T_2 são estimadores centrados para o parâmetro θ , com $\theta > 0$. Então, pode-se concluir que:

T_1 e T_2 são consistentes	
$E(T_1) = E(T_2) = 0$	
$Var(T_1) = Var(T_2)$	
se T_1 é mais eficiente que T_2 , então $EQM(T_1) \leq EQM(T_2)$	

4. [20] Seja (X_1, X_2, \dots, X_n) uma amostra casual simples retirada de uma população X com média $E(X) = 2\theta$, variância $Var(X) = 2\theta^2$ e função de densidade:

$$f(x; \theta) = \frac{x}{\theta^2} \exp\left(-\frac{x}{\theta}\right), \quad x > 0, \quad \theta > 0.$$

Mostre que os estimadores da máxima verosimilhança e do método dos momentos para o parâmetro θ são iguais, e estude a sua consistência.

6. Com base numa amostra casual de 308 empresas, foi estimado o seguinte modelo de regressão linear múltipla, com o objectivo de estudar o nível de investimento em investigação e desenvolvimento:

$$\log(\text{RD}_i) = \beta_0 + \beta_1 \text{ATIV}_i + \beta_2 \text{TECMIX}_i + \beta_3 \text{FUNC.G30}_i + \beta_4 \text{FUNC.LEQ30}_i + u_i$$

onde \log representa o logaritmo natural, e as variáveis têm o seguinte significado:

RD: montante investido anualmente em investigação e desenvolvimento (milhares de Euros);

ATIV: número de anos de atividade da empresa;

TECMIX: índice do nível tecnológico da empresa (valores de 1 = baixo a 6 = elevado);

FUNC.G30: número de funcionários com idade superior a 30 anos;

FUNC.LEQ30: número de funcionários com idade inferior ou igual a 30 anos.

Tendo em conta os resultados da estimação do modelo na Equação 1 e das restantes regressões auxiliares apresentadas no Anexo, responda as seguintes questões:

- a) [15] Supondo satisfeitas as hipóteses do modelo de regressão linear, teste ao nível de 5% a significância estatística individual dos parâmetros β_1 e β_2 da **Equação 1** e interprete as suas estimativas.

- b) Pelos resultados da **Equação 2** podemos concluir que

não há evidência de heterocedasticidade nos erros do modelo da Equação 1	
a forma funcional apresenta erros de especificação	
não há evidência de má especificação na forma funcional	
a variância dos erros depende negativamente do número de funcionários com mais de 30 anos	

c) [20] Será que a evidência estatística é favorável à hipótese de que o efeito de mais um funcionário com idade superior a 30 anos é compensado por mais um funcionário com idade inferior ou igual a 30 anos? Justifique através de um teste estatístico adequado.

d) [15] Avalie a hipótese de que o número de funcionários da empresa não influencia a quantidade investida anualmente em investigação e desenvolvimento. Justifique através de um teste de hipótese adequado.

- e) Uma empresa acaba de contratar 3 novos funcionários, dos quais apenas 1 tem idade inferior ou igual a 30 anos. Segundo a **Equação 1**, qual é aproximadamente a alteração percentual esperada no valor do investimento em investigação e desenvolvimento, mantendo tudo o resto constante?

4.3%	
7.0%	
1.7%	
10.4%	

- f) [20] Um investigador suspeita que o número de anos de atividade possa ter um efeito quadrático. Se isso for verdade, qual as consequências sobre os estimadores OLS dos parâmetros da **Equação 1**? O investigador reestimou o modelo, acrescentando como regressor adicional $YFIT^2$, onde $YFIT$ são os valores ajustados da **Equação 1**. Sabendo que obteve para o correspondente coeficiente um valor de -1.40115 , com standard error igual a 0.892599 , será que podemos concluir que o investigador tinha razão?

Anexo EN Janeiro 2017

Equação 1

Dependent Variable: log(RD)
Included observations: 308

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	6.468145	0.168897	38.30	0.000
ATIV	0.005876	0.002257	2.60	0.010
TECMIX	-0.039838	0.014879	-2.68	0.008
FUNC.G30	0.089042	0.025668	3.47	0.001
FUNC.LEQ30	-0.108417	0.058574	-1.85	0.065
R-squared	0.0817	Sum squared resid		100.1171
Adjusted R-squared	0.0696	F-statistic		6.74
		Prob(F-statistic)		0.000

Equação 2

Dependent Variable: RESID^2
Included observations: 308

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.306369	0.119608	2.56	0.011
ATIV	0.000125	0.001599	0.08	0.938
TECMIX	0.009147	0.010537	0.87	0.386
FUNC.G30	-0.019068	0.018178	-1.05	0.295
FUNC.LEQ30	0.034835	0.041480	0.84	0.402
R-squared	0.0069	Sum squared resid		50.20971
Adjusted R-squared	0.0063	F-statistic		0.52
		Prob(F-statistic)		0.719

Nota: a variável RESID representa os resíduos da Equação 1

Equação 3

Dependent Variable: log(RD)
Included observations: 308

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	6.468145	0.168897	38.30	0.000
ATIV	0.005876	0.002257	2.60	0.010
TECMIX	-0.039838	0.014879	-2.68	0.008
FUNC.G30 - FUNC.LEQ30	0.089042	0.025668	3.47	0.001
FUNC.LEQ30	-0.019375	0.049590	-0.39	0.696
R-squared	0.0817	Sum squared resid		100.1171
Adjusted R-squared	0.0696	F-statistic		6.74
		Prob(F-statistic)		0.000

Equação 4

Dependent Variable: log(RD)
Included observations: 308

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	6.752736	0.142166	47.5	0.000
ATIV	0.004994	0.002166	2.31	0.022
TECMIX	-0.043528	0.015082	-2.89	0.004
R-squared	0.0452	Sum squared resid		104.0938
Adjusted R-squared	0.0389	F-statistic		7.22
		Prob(F-statistic)		0.000