Nome: ______________________________________ Nº__________

Espaço reservado a classificações				
1.a)	1.b)		1.a)	1.b)
1.c)	1.d)		1.c)	1.d)
2.a)	2.b)	2.c)	1.e)	1.f)
PARTE I:			**PARTE II:**	

Se necessitar de mais espaço, pode utilizar a última página do enunciado, indicando claramente a respectiva questão.
É expressamente proibido destacar as folhas do enunciado!

Boa Sorte!

PARTE II

1) Um estudo econométrico de regressão linear levado a cabo sobre as empresas tecnológicas portuguesas baseou-se no seguinte modelo:

$$RESULT_t = \beta_1 + \beta_2 CAP_t + \beta_3 PESS_t + \beta_4 IMP_t + \beta_5 NEMP_t + u_t$$

Onde:

- $RESULT_t$ – resultado líquido da empresa t (em milhares de euros);
- CAP_t – capital próprio da empresa t (em milhares de euros);
- $PESS_t$ – custos com pessoal da empresa t (em milhares de euros);
- IMP_t – impostos pagos pela empresa t (em milhares de euros);
- $NEMP_t$ – número de empregados da empresa t.

No anexo encontram-se modelos de regressão estimados que deve utilizar para responder às questões que se seguem.

a) Interprete as estimativas dos coeficientes associados com os regressores $\boldsymbol{CAP}$ e $\boldsymbol{NEMP}$ na **Equação 1**. *(1.5)*

$b_2 = 0.032 \rightarrow$ *Ceteris paribus*, estima-se em média, que se o capital próprio da empresa aumentar em 1 milhar de euros, o resultado líquido da empresa aumenta 32 euros.

$b_5 = -7.167 \rightarrow$ *Ceteris paribus*, estima-se em média, que por cada empregado adicional na empresa, o resultado líquido da empresa diminui 7167 euros.

b) Teste ao nível de 5% a significância individual dos regressores $\boldsymbol{CAP}$ e $\boldsymbol{NEMP}$ na **Equação 1**. *(1.5)*

Significância individual de *CAP:*

$$H_0: \beta_2 = 0 \quad \text{vs} \quad H_1: \beta_2 \neq 0$$

$$\text{ET: } t_2 = \frac{b_2}{s_{b_2}} \sim t(35)$$

Do output obtém-se $p_{2,obs} = 0.287$. Porque $p_{2,obs} > \alpha = 0.05$, não se rejeita H_0 ao nível de 5%. Logo, o regressor *CAP* não é estatisticamente significativo.

Significância individual de *NEMP:*

$$H_0: \beta_5 = 0 \quad \text{vs} \quad H_1: \beta_5 \neq 0$$

$$\text{ET: } t_5 = \frac{b_5}{s_{b_5}} \sim t(35)$$

Do output obtém-se $p_{5,obs} = 0.001$. Porque $p_{5,obs} < \alpha = 0.05$, rejeita-se H_0 ao nível de 5%. Logo, o regressor *NEMP* é estatisticamente significativo.

c) Teste a significância global da **Equação 1** ao nível de 5%. *(1.5)*

$$H_0: \beta_2 = \beta_3 = \beta_4 = \beta_5 = 0 \quad \text{vs} \quad H_1: \exists \beta_j \neq 0 \ (j = 2, 3, 4, 5)$$

$$\text{ET: } F = \frac{R^2/(k-1)}{(1-R^2)/(n-k)} \sim F(4, 35)$$

Do output, obtém-se: $F_{obs} = 21.958$.

$W_{0.05} = \{f_{obs}: f_{obs} > f_{0.05}\}$, onde $2.61 < f_{0.05} < 2.69$

Porque $f_{obs} \in W_{0.05}$, rejeita-se H_0 ao nível de 5%. Logo, conclui-se que a Equação 1 é globalmente significativa.

d) As variáveis $\boldsymbol{IMP}$ e $\boldsymbol{NEMP}$ são, no seu conjunto, relevantes para explicar o resultado líquido? Justifique com base num teste adequado ao nível de 5%. *(1.5)*

$$H_0: \beta_4 = \beta_5 = 0 \quad \text{vs} \quad H_1: \exists \beta_j \neq 0 \; (j = 4, 5)$$

$$\text{ET: } F = \frac{(R^2 - R_0^2)/m}{(1 - R^2)/(n - k)} \sim F(2, 35)$$

$W_{0.05} = \{f_{obs}: f_{obs} > f_{0.05}\}$, onde $3.23 < f_{0.05} < 3.32$

$$F_{obs} = \frac{(0.715 - 0.160)/2}{(1 - 0.715)/35} \approx 34.079$$

Porque $f_{obs} \in W_{0.05}$, rejeita-se H_0 ao nível de 5%. Logo, conclui-se que as variáveis *IMP* e *NEMP* são conjuntamente significativas.

e) Indique como poderia detectar a presença de heterocedasticidade na **Equação 2**. Formalize o problema em termos das hipóteses a testar, estatística de teste e respectiva distribuição. Recomende uma possível medida correctiva no caso de ser detectada heterocedasticidade. *(2.0)*

Uma maneira de averiguar a presença de heterocedasticidade seria efectuar o teste de White.
Optando pela versão simplificada, teríamos de estimar a equação auxiliar de teste:

$$\hat{u}_t^2 = \alpha_1 + \alpha_2 \hat{y}_t + \alpha_3 \hat{y}_t^2 + v_t,$$

onde $\hat{u}_t$ e $\hat{y}_t$ são, respectivamente, os resíduos e os valores ajustados, obtidos da estimação da Equação 2.
Assim, deve testar-se na regressão auxiliar:

$$H_0: \alpha_2 = \alpha_3 = 0 \quad \text{vs} \quad H_1: \exists \alpha_j \neq 0 \; (j = 2, 3)$$

$$\text{ET: } W = nR_a^2 \overset{a}{\sim} \chi^2(2),$$

onde R_a^2 é o coeficiente de determinação da regressão auxiliar.
Caso se rejeitasse H_0, teríamos evidência de heterocedasticidade, e uma possível medida correctiva seria estimar a matriz das covariâncias, $Cov(b|X)$, utilizando o estimador robusto à heterocedasticidade de White.

f) Defina uma variável binária indicativa do grau de internacionalização da empresa (1, se "multinacional"; 0 se "nacional"). Como poderia testar a alteração de estrutura na **Equação 2** entre as empresas nacionais e multinacionais usando a variável binária? Formalize o problema em termos das hipóteses a testar, estatística de teste e respectiva distribuição. *(2.0)*

Seja a variável binária,

$$MULT_t = \begin{cases} 1, \text{se empresa } t \text{ é multinacional} \\ 0, \text{se empresa } t \text{ é nacional} \end{cases}$$

Para detectar uma possível alteração de estrutura na Equação 2 entre as empresas multinacionais e as empresas nacionais, teríamos de estimar a seguinte regressão aumentada:

$$RESULT_t = \beta_1 + \delta_1 MULT_t + \beta_2 CAP_t + \delta_2 CAP_t \times MULT_t + \beta_3 PESS_t + \delta_3 PESS_t \times MULT_t + e_t$$

E efectuar o seguinte teste de hipóteses:

$$H_0: \delta_1 = \delta_2 = \delta_3 = 0 \quad \text{vs} \quad H_1: \exists \delta_j \neq 0 \ (j = 1, 2, 3)$$

$$\text{ET: } F = \frac{(R^2 - R_0^2)/3}{(1 - R^2)/(40 - 6)} \sim F(3, 34)$$

Onde R^2 é o coeficiente de determinação desta regressão aumentada e R_0^2 é o coeficiente de determinação da Equação 2.

Caso se rejeitasse H_0, teríamos evidência de alteração de estrutura na Equação 2, o que implicaria a existência de modelos distintos para explicar o resultado líquido para empresas multinacionais e empresas nacionais.

PARTE I

1. Os distribuidores de pizas ao domicílio de uma certa pizaria recebem por vezes gorjetas dos clientes. Como tal, o dono dessa pizaria está interessado em estimar a média das gorjetas recebidas. Assuma que o montante, em euros, de cada gorjeta é uma variável aleatória X com média μ e variância σ^2 desconhecidas. Considere ainda que para uma amostra casual de $n = 100$ entregas ao domicílio se obteve $\bar{x} = 1.75$ e $s'^2 = 1$.

   a) Mostre que os estimadores do método dos momentos para os parâmetros desconhecidos são: $\tilde{\mu} = \bar{X}$ e $\tilde{\sigma}^2 = S^2 = \left(\sum_{i=1}^{n} X_i^2\right)/n - \bar{X}^2$. *(2.0)*

$$\begin{cases} E(X) = \bar{X} \\ E(X^2) = \dfrac{\sum_{i=1}^{n} X_i^2}{n} \end{cases} \Leftrightarrow \begin{cases} \tilde{\mu} = \bar{X} \\ Var(X) + E(X)^2 = \dfrac{\sum_{i=1}^{n} X_i^2}{n} \end{cases} \Leftrightarrow \begin{cases} \tilde{\mu} = \bar{X} \\ \sigma^2 + \mu^2 = \dfrac{\sum_{i=1}^{n} X_i^2}{n} \end{cases} \Leftrightarrow \begin{cases} \tilde{\mu} = \bar{X} \\ \sigma^2 = \dfrac{\sum_{i=1}^{n} X_i^2}{n} - \mu^2 \end{cases} \Leftrightarrow$$

$$\Leftrightarrow \begin{cases} \tilde{\mu} = \bar{X} \\ \tilde{\sigma}^2 = \dfrac{\sum_{i=1}^{n} X_i^2}{n} - \tilde{\mu}^2 \end{cases} \Leftrightarrow \begin{cases} \tilde{\mu} = \bar{X} \\ \tilde{\sigma}^2 = \dfrac{\sum_{i=1}^{n} X_i^2}{n} - \bar{X}^2 = S^2 \end{cases}$$

   b) As respectivas estimativas do método dos momentos para os parâmetros são:
   (Resposta certa: 1 / Resposta errada: -0.25)

$\tilde{\mu} = 1.75$ e $\tilde{\sigma}^2 = 0.99^2$	
$\tilde{\mu} = 1.75$ e $\tilde{\sigma}^2 = 0.99$	**X**
$\tilde{\mu} = 1.75$ e $\tilde{\sigma}^2 = \sqrt{0.99}$	

c) Construa um intervalo de confiança a 99% para a média das gorjetas μ, e interprete o resultado obtido arredondado às décimas. *(1.5)*

<u>IC a 99% para μ (Grandes Amostras: Caso Geral)</u>

$$\text{VF: } Z = \frac{\bar{X} - \mu}{S'/\sqrt{n}} \overset{a}{\sim} N(0,1)$$

IC: $\left(\bar{x} \pm z_{0.005} \times s'/\sqrt{n}\right) = \left(1.75 \pm 2.576 \times 1/\sqrt{100}\right) = (1.4924, 2.0076)$

Com uma confiança de 99%, podemos concluir que a média das gorjetas dos clientes se situa aproximadamente entre 1.5 e 2 euros.

d) Após alguma reflexão, o dono da pizaria concluiu que a média das gorjetas não chega a 2 euros. Comente esta afirmação do dono da pizaria efectuando um teste de hipóteses adequado com um nível de 5%. *(1.5)*

$$H_0: \mu \geq 2 \quad \text{vs} \quad H_1: \mu < 2$$

$$\text{ET: } Z = \frac{\bar{X} - 2}{S'/\sqrt{n}} \overset{a}{\sim} N(0,1)$$

$W_{0.05} = \{z_{obs}: z_{obs} < -1.645\}$

$$z_{obs} = \frac{1.75 - 2}{1/\sqrt{100}} = -2.5$$

Porque $z_{obs} \in W_{0.05}$, rejeita-se H_0 ao nível de 5%. A evidência estatística é favorável à afirmação do dono da pizaria.

2. Ainda de acordo com o exercício anterior, o dono da pizaria está agora interessado na frequência θ com que os clientes dão gorjeta. Assim sendo, definiu uma variável aleatória de Bernoulli, Y, como sendo o indicador da gorjeta, isto é, $Y = 1$ se o distribuidor recebeu gorjeta e $Y = 0$ caso contrário. Usando a mesma amostra casual de $n = 100$ entregas ao domicílio, o dono da pizaria obteve $\sum_{i=1}^{100} y_i = 41$.

a) Sabendo que o estimador da máxima verosimilhança é $\hat{\theta} = \bar{Y}$, mostre que este é consistente para θ. *(1.5)*

<u>População de Bernoulli:</u> $Y \sim B(1, \theta)$

$$\lim_{n\to\infty} E(\hat{\theta}) = \lim_{n\to\infty} E(\bar{Y}) = \lim_{n\to\infty} E(Y) = \lim_{n\to\infty} \theta = \theta$$

$$\lim_{n\to\infty} Var(\hat{\theta}) = \lim_{n\to\infty} Var(\bar{Y}) = \lim_{n\to\infty} \frac{Var(Y)}{n} = \lim_{n\to\infty} \frac{\theta(1-\theta)}{n} = 0$$

Porque $\lim_{n\to\infty} E(\hat{\theta}) = \theta$ e $\lim_{n\to\infty} Var(\hat{\theta}) = 0$, conclui-se que o estimador da máxima verosimilhança $\hat{\theta} = \bar{Y}$ é consistente para θ.

b) A percentagem de clientes que não dão gorjeta aos distribuidores é dada pela expressão: $100(1-\theta)$. Calcule, justificando, a estimativa da máxima verosimilhança para esta percentagem. *(1.5)*

Quer-se obter a estimativa de máxima verosimilhança para $100(1-\theta)$. Porque a expressão $100(1-\theta)$ é uma função biunívoca de θ, pode usar-se a propriedade da invariância dos estimadores de máxima verosimilhança, e assim obter:

$$100\widehat{(1-\theta)} = 100(1-\hat{\theta}) = 100(1-\bar{y}) = 100(1-41/100) = 100(1-0.41) = 100 \times 0.59 = 59$$

Logo, estima-se que a percentagem de clientes que não dão gorjeta é de 59%.

c) O intervalo de confiança a 95% para θ é: $(0.3136, 0.5064)$. Então:
(Resposta certa: 1 / Resposta errada: -0.25)

Rejeita-se $H_0: \theta = 0.5$ contra $H_1: \theta \neq 0.5$ para $\alpha = 0.05$.	
Não se rejeita $H_0: \theta = 0.5$ contra $H_1: \theta \neq 0.5$ para $\alpha = 0.10$.	
Rejeita-se $H_0: \theta = 0.5$ contra $H_1: \theta \neq 0.5$ para $\alpha = 0.025$.	
Não se rejeita $H_0: \theta = 0.5$ contra $H_1: \theta \neq 0.5$ para $\alpha = 0.05$.	**X**

ANEXO

Equação 1: $RESULT_t = \beta_1 + \beta_2 CAP_t + \beta_3 PESS_t + \beta_4 IMP_t + \beta_5 NEMP_t + u_t$

Regression Statistics	
Multiple R	0.846
R Square	0.715
Adjusted R Square	0.682
Standard Error	3751.874
Observations	40

ANOVA

	df	*SS*	*MS*	*F*
Regression	4	1236393527	309098381.8	21.958
Residual	35	492679686.2	14076562.46	
Total	39	1729073214		

	Coefficients	*Standard Error*	*t Stat*	*P-value*
Intercept	697.852	802.648	0.869	0.391
CAP	0.032	0.029	1.081	0.287
PESS	0.145	0.072	2.006	0.053
IMP	2.154	0.305	7.056	0.000
NEMP	-7.167	2.026	-3.537	0.001

Equação 2: $RESULT_t = \beta_1 + \beta_2 CAP_t + \beta_3 PESS_t + u_t$

Regression Statistics	
Multiple R	0.400
R Square	0.160
Adjusted R Square	0.115
Standard Error	6264.437
Observations	40

ANOVA

	df	*SS*	*MS*	*F*
Regression	2	277075915.6	138537957.8	3.530
Residual	37	1451997298	39243170.21	
Total	39	1729073214		

	Coefficients	*Standard Error*	*t Stat*	*P-value*
Intercept	950.916	1321.184	0.720	0.476
CAP	0.101	0.046	2.171	0.036
PESS	-0.026	0.076	-0.342	0.734