

Nome: \_\_\_\_\_ Número: \_\_\_\_\_

**Notas:** Certifique-se de que o seu telemóvel está desligado. Se não estiver é motivo suficiente para anulação da prova. As perguntas de escolha múltipla valem 1 valor; respostas erradas são penalizadas em 0.25. Caso nada seja dito em contrário utilize um nível de significância de 5% nos testes de hipóteses que efetuar. Fundamente e formalize devidamente todas as respostas. Pode usar a última página para continuar qualquer questão.

Espaço reservado para classificações				
1)	2)	6)	7)	EM(3, 4, 5):
8a)	8c)	8e)	8g)	EM(8.b, d, f):

1. O *reality show* “A Horta” tem grande sucesso entre os telespectadores. Para estudar as audiências recolheu-se, de forma independente, informação junto de homens e mulheres. De 100 homens inquiridos, 70 responderam ver o programa, enquanto num total de 150 mulheres, 114 responderam igualmente ser fãs do programa. De acordo com um intervalo de confiança a 99% pode afirmar que, na população, a proporção de mulheres que vê o programa é superior à proporção de homens? [1.5]

2. O gerente de um *resort* afirma que o preço médio de uma estadia não é superior a 500 euros. Suponha que o preço de uma estadia é uma variável aleatória com desvio padrão igual a 40 euros. Para efetuar o teste proposto rejeita-se a afirmação do gerente quando a média da amostra for superior a 510 euros. Considerando uma amostra casual de 50 estadias, qual é aproximadamente o nível de significância associado a este teste? Não se esqueça de definir as hipóteses em teste. [2.0]

3. Assinale a alternativa correta:

- Todas as variáveis fulcrais são estimadores.
- Todos os estimadores são variáveis fulcrais.
- Todos os estimadores são estatísticas.
- Todas as anteriores.

4. Se um teste de hipótese simples contra uma alternativa simples tem potência 0.8, isso significa que:

- quando  $H_0$  é falsa, a probabilidade de  $H_0$  ser rejeitada é de 0.8.
- quando  $H_0$  é verdadeira, a probabilidade de  $H_0$  ser rejeitada é de 0.8.
- quando  $H_0$  é verdadeira, a probabilidade de  $H_0$  não ser rejeitada é de 0.8.
- quando  $H_0$  é falsa, a probabilidade de  $H_0$  não ser rejeitada é de 0.8.

5. Assinale a alternativa correta.

Os reagrupamentos de classes num teste de ajustamento...

- ... devem ser feitos com base nas frequências observadas.
- ... devem ser feitos com base nas frequências esperadas.
- ... devem ser feitos quando o número de classes é par.
- ... devem ser feitos quando o número de classes é ímpar.

6. Represente-se a duração de um anúncio publicitário, em segundos, por uma variável aleatória  $X$ . Para testar se  $X$  tem distribuição exponencial de média 10 segundos, recolheu-se uma amostra de 100 anúncios, registando-se a seguinte informação:

Duração (Segundos)	(0, 5]	(5, 10]	(10, 15]	(15, 20]	$\geq 20$
Frequência Observada	25	30	15	20	10
Frequência Esperada	<b>A</b>	<b>B</b>	14.475	8.779	13.534

Calcule as duas primeiras frequências esperadas em falta no quadro (**A** e **B**) e efetue o teste em questão, considerando um nível de 1%. Que pode concluir? [2.0]

7. Admita que numa determinada região o número de vezes que uma pessoa vai ao ginásio por semana é uma variável aleatória  $X$  com a seguinte função probabilidade:

$$f(x|\theta) = \frac{\theta^x}{(\theta + 1)^{x+1}} \quad (x = 0, 1, 2, \dots; \theta > 0)$$

Recolhida uma amostra casual de 10 pessoas desta região obteve-se  $\sum_{i=1}^{10} x_i = 6$ . Sabendo que o estimador da máxima verosimilhança para  $\theta$  é  $\bar{X}$  calcule, justificando, a estimativa para a percentagem de pessoas desta região que não frequenta o ginásio. [1.5]

8. O estudo econométrico que se segue baseia-se na análise ao consumo mensal de internet em função de um conjunto de características pessoais. Para tal, foi definido o seguinte modelo de regressão linear:

$$int = \beta_0 + \beta_1 \log(rend) + \beta_2 ncontas + \beta_3 idade + \beta_4 idade^2 + u$$

Onde:

- *int* – consumo mensal de internet do indivíduo, em gigabytes;
- *rend* – rendimento mensal do indivíduo, em euros;
- *ncontas* – número de contas ativas do indivíduo em redes sociais;
- *idade* – idade do indivíduo, em anos.

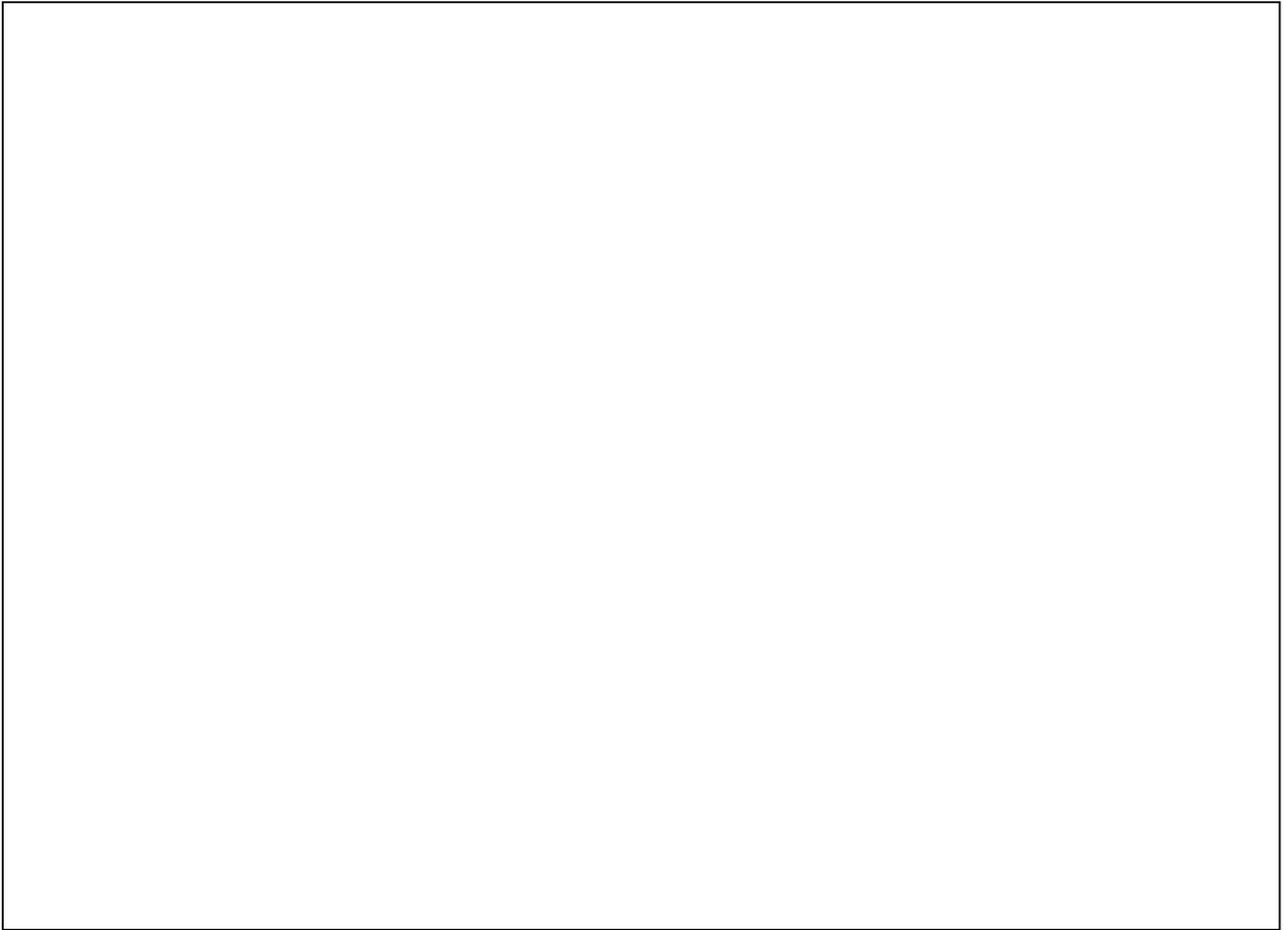
A estimação do modelo inicial encontra-se na Equação 1, em Anexo. Igualmente no Anexo encontrará outras regressões estimadas que lhe serão úteis para as questões que se seguem.

a) Interprete a estimativa associada ao regressor “ $\log(rend)$ ” e teste a significância estatística deste. [1.5]

b) O valor do coeficiente de determinação da Equação 1 significa que:

- 29.75% dos resíduos são iguais a zero.
- o estimador OLS apenas utilizou 29.75% da amostra.
- a probabilidade de o modelo estar errado é igual a 0.2975.
- os regressores explicam 29.75% da variação total amostral do consumo de internet.

c) Com base num teste de hipóteses adequado verifique se a idade da pessoa é um fator estatisticamente relevante no consumo de internet. [1.5]



d) Qual é aproximadamente a idade da pessoa a partir da qual o efeito sobre o consumo de internet passa a ser negativo?

- 25 anos
- 30 anos
- 35 anos
- 40 anos

e) Construa um intervalo de confiança a 95% para  $\theta = \frac{\beta_1}{100} + \beta_2$  e interprete o resultado. [2.0]

f) Escolha a alternativa correta.

De acordo com o seguinte output do *EViews*, pode concluir-se que:

Ramsey RESET Test  
Equation: EQ01  
Specification: INT C LOG(REND) NCONTAS IDADE IDADE^2

	Value	df	Probability
F-statistic	9.876384	(2, 493)	0.0001

- A probabilidade de o modelo inicial estar bem especificado é igual a 0.0001.
- Existe evidência estatística de que o modelo inicial sofre de má especificação.
- Existe evidência estatística de que os erros do modelo inicial são homocedásticos.
- Existe evidência estatística de que o modelo inicial está bem especificado.

g) Diga qual o objetivo da estimação da Equação 4 em Anexo e tire as principais conclusões a ela associadas no que respeita às propriedades do OLS. **[2.0]**

CONTINUAÇÃO DA QUESTÃO \_\_\_\_\_



## ANEXO

### Equação 1

Dependent Variable: INT  
Method: Least Squares  
Included observations: 500

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-185.6037	62.78577	-2.956143	0.0033
LOG(REND)	14.88069	1.475544	10.08488	0.0000
NCONTAS	0.565590	0.225524	2.507892	0.0125
IDADE	6.319112	3.690368	1.712326	0.0875
IDADE^2	-0.078814	0.054553	-1.444716	0.1492
R-squared	0.297533	Mean dependent var		54.99000
Adjusted R-squared	0.291857	S.D. dependent var		12.54051
S.E. of regression	10.55300	Akaike info criterion		7.560646
Sum squared resid	55126.04	Schwarz criterion		7.602793
Log likelihood	-1885.162	Hannan-Quinn criter.		7.577185
F-statistic	52.41492	Durbin-Watson stat		1.828068
Prob(F-statistic)	0.000000			

### Equação 2

Dependent Variable: INT  
Method: Least Squares  
Included observations: 500

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-81.41585	10.52907	-7.732481	0.0000
LOG(REND)	17.23404	1.414457	12.18421	0.0000
NCONTAS	0.685495	0.227860	3.008400	0.0028
R-squared	0.264774	Mean dependent var		54.99000
Adjusted R-squared	0.261815	S.D. dependent var		12.54051
S.E. of regression	10.77452	Akaike info criterion		7.598226
Sum squared resid	57696.82	Schwarz criterion		7.623514
Log likelihood	-1896.557	Hannan-Quinn criter.		7.608149
F-statistic	89.49135	Durbin-Watson stat		1.848178
Prob(F-statistic)	0.000000			

### Equação 3

Dependent Variable: INT  
 Method: Least Squares  
 Included observations: 500

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-185.6037	62.78577	-2.956143	0.0033
LOG(REND)-NCONTAS/100	14.88069	1.475544	10.08488	0.0000
NCONTAS	0.714397	0.223958	3.189872	0.0015
IDADE	6.319112	3.690368	1.712326	0.0875
IDADE^2	-0.078814	0.054553	-1.444716	0.1492
R-squared	0.297533	Mean dependent var		54.99000
Adjusted R-squared	0.291857	S.D. dependent var		12.54051
S.E. of regression	10.55300	Akaike info criterion		7.560646
Sum squared resid	55126.04	Schwarz criterion		7.602793
Log likelihood	-1885.162	Hannan-Quinn criter.		7.577185
F-statistic	52.41492	Durbin-Watson stat		1.828068
Prob(F-statistic)	0.000000			

### Equação 4

Dependent Variable: RES^2  
 Method: Least Squares  
 Included observations: 500

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	871.2516	300.0341	2.903842	0.0039
FITTED	-25.88942	11.51271	-2.248768	0.0250
FITTED^2	0.215808	0.109702	1.967221	0.0497
R-squared	0.023825	Mean dependent var		110.2521
Adjusted R-squared	0.019897	S.D. dependent var		180.3533
S.E. of regression	178.5500	Akaike info criterion		13.21360
Sum squared resid	15844413	Schwarz criterion		13.23888
Log likelihood	-3300.399	Hannan-Quinn criter.		13.22352
F-statistic	6.065098	Durbin-Watson stat		1.785897
Prob(F-statistic)	0.002498			

**NOTA: As variáveis RES e FITTED são os resíduos e os valores ajustados da Equação 1, respetivamente.**