

**Instruções:**

1. Verifique se o seu telemóvel está desligado. Ter o telemóvel ligado durante a prova **é motivo para anulação** da mesma.
2. **Formalize e fundamente** as suas respostas.
3. Durante o decorrer da prova não serão prestados quaisquer esclarecimentos. Se tiver alguma dúvida pode apresentá-la por escrito.
4. Caso não seja dito nada em contrário utilize uma dimensão de 5% nos testes estatísticos que efectuar.
5. Responda à questão **6** em **folha separada** das questões anteriores.

(10) 1. Considere uma variável aleatória  $X$  com função de distribuição que depende do parâmetro  $\alpha$  e que verifica  $E(X) = (1 - \alpha)/\alpha$ . Tendo dados de uma amostra aleatória de 150 observações para a qual se obteve:  $\sum_{i=1}^{150} x_i = 147$  determine uma estimativa para  $\alpha$ .

(10) 2. Concorda com a seguinte afirmação: “O desvio padrão da média da amostra tenderá a ser maior que o desvio padrão da média da população”? Justifique sucintamente.

(20) 3. Uma sondagem a 500 eleitores escolhidos aleatoriamente deu uma intenção de voto no partido A nas próximas eleições legislativas de 35.4%. Depois de ter ocorrido um debate na TV entre o líder deste partido e o seu maior rival foi realizada nova sondagem a 600 eleitores escolhidos aleatoriamente tendo-se observado um aumento de 3.4 pontos percentuais nas intenções de voto em A. Um comentador afirmou que este resultado indica que o partido A aumentou a sua proporção de votantes na população depois do debate. Concorda com o comentador? Justifique através de um teste de hipóteses.

(20) 4. Com o objectivo de testar se as classificações obtidas num exame de Estatística têm distribuição Normal observou-se uma amostra aleatória com 82 alunos obtendo-se os seguintes resultados,

$x$ - classificação	$x \leq 5$	$5 < x \leq 9$	$9 < x \leq 12$	$12 < x \leq 15$	$x > 15$	$\bar{x} = 9.82$
Nº de alunos	22	15	17	15	13	$s' = 5$

Sabe-se que o valor obtido para a estatística de teste é igual a 7.44. Calcule a parcela desta estatística **apenas para a classe**  $9 < x \leq 12$ . Se a hipótese de que as notas têm distribuição normal for verdadeira, o valor 7.44 é uma concretização de uma variável aleatória com que distribuição? Que pode concluir?

5. A quantidade de poluente emitida por uma empresa por cada metro cúbico ( $m^3$ ) de água de uma barragem é uma variável com distribuição desconhecida. A lei prevê uma multa no caso da média da quantidade de poluente em cada  $m^3$  de toda a barragem exceder 30 gramas ( $\mu > 30$ ). Mas, dada a impossibilidade de analisar toda a barragem, a lei prevê a deliberação com base numa amostra de  $50 m^3$  de água (retirados em 50 porções aleatórias de  $1m^3$  do total da barragem). Sabe-se que o desvio padrão corrigido da **média** da amostra é igual 1.5 gramas.

(20) a) Pretendendo-se que a probabilidade de condenar indevidamente uma empresa seja de 0.1 determine a partir de que valor da média da amostra se deve condenar a empresa.

(20) b) Suponha que a lei definiu que, se a média de poluente da amostra for superior a 32 gramas, a empresa será condenada. Qual a probabilidade de condenar uma empresa que lança uma média de 35 gramas de poluente por metro cúbico de água em toda a barragem? E se essa média for igual a 40, a probabilidade será maior ou menor? Justifique sem fazer cálculos.

6. Para explicar a quantidade de maçãs biológicas consumidas semanalmente por uma família especificou-se o seguinte modelo,

$$MACBIO = \beta_0 + \beta_1 PMACBIO + \beta_2 PMACNORM + \beta_3 FAMREND + \beta_4 EDUC + \beta_5 NAGFAM + u$$

onde as variáveis têm o seguinte significado,

<i>MACBIO</i>	- consumo semanal de maçãs biológicas em kg da família,
<i>PMACBIO</i>	- preço médio de 1kg de maçãs biológicas em euros,
<i>PMACNORM</i>	- preço médio de 1kg de maçãs normais em euros,
<i>FAMREND</i>	- rendimento médio anual do agregado familiar em milhares de euros,
<i>EDUC</i>	- Escolaridade média dos adultos do agregado familiar,
<i>NAGFAM</i>	- número de pessoas do agregado familiar.

Na Equação 1 do Anexo encontra-se a estimação deste modelo e ainda outras estimações que permitem responder às várias questões deste problema.

- (20) a) Interprete as estimativas obtidas respectivamente para  $\beta_1$  e  $\beta_5$  na Equação 1 e avalie se o sinal das estimativas obtidas para todos os coeficientes faz sentido do ponto de vista económico. Que pode concluir a partir do resultado obtido para o R-squared desta Equação?
- (20) b) Será que a evidência estatística é favorável à hipótese de que um aumento de um euro do preço das maçãs biológicas é compensado pelo aumento de um euro do preço das outras maçãs? Justifique através de um teste estatístico adequado.
- (20) c) Sabendo que a Equação 2 é uma regressão auxiliar de um teste RESET escreva a expressão que permite obter a variável FIT e indique o seu significado. Que conclusão pode retirar a partir dos resultados desta equação?
- (20) d) Faça a previsão pontual e por intervalo de confiança a 95% da média do consumo semanal de maçãs biológicas das famílias com 4 pessoas, um rendimento médio anual de 60 000€ e uma escolaridade média dos seus adultos de 15 anos, quando o preço por kg destas maçãs é de 2.5€ enquanto das maçãs normais é de 1.5€
- (10) e) Concorda com a seguinte afirmação: “Não faz sentido testar  $\beta_3 = 0$  pois obteve-se  $\hat{\beta}_3 = 0.003$  que é diferente de zero e portanto pode-se concluir logo que  $\beta_3 \neq 0$ ”? Justifique sucintamente.
- (10) f) Considere a equação,

$$y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + u \quad (1)$$

Suponha que  $\beta_2 \neq 0$ . Para estimar o efeito parcial de  $X_1$  sobre  $y$ ,  $\beta_1$ , um aluno afirma que basta usar o estimador OLS numa regressão de  $y$  sobre  $X_1$  (omitindo  $X_2$ ). Concorda com este aluno? Justifique sucintamente.

# ANEXO

## Equação 1

Dependent Variable: MACBIO

Method: Least Squares

Included observations: 660

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	1.363146	0.155418	8.770838	0.0000
PMACBIO	-0.146821	0.006650	-22.07964	0.0000
PMACNORM	0.164934	0.079270	2.080645	0.0379
FAMREND	0.002768	0.000570	4.852339	0.0000
EDUC	0.028262	0.008951	3.157377	0.0017
NAGFAM	0.086866	0.012792	6.790782	0.0000
R-squared	0.468827	Mean dependent var		2.130228
Adjusted R-squared	0.464766	S.D. dependent var		0.676247
S.E. of regression	0.494740	Akaike info criterion		1.439481
Sum squared resid	160.0782	Schwarz criterion		1.480320
F-statistic	115.4474	Durbin-Watson stat		1.903987
Prob(F-statistic)	0.000000			

### Matriz de variâncias-covariâncias

	C	PMACBIO	PMACNORM	FAMREND	EDUC	NAGFAM
C	0.024155	-9.26E-05	-0.006452	7.39E-06	-0.001153	-0.000607
PMACBIO	-9.26E-05	4.42E-05	3.06E-05	-1.24E-07	1.92E-06	-4.12E-06
PMACNORM	-0.006452	3.06E-05	0.006284	7.68E-07	5.34E-05	1.94E-05
FAMREND	7.39E-06	-1.24E-07	7.68E-07	3.25E-07	-1.56E-06	-9.78E-07
EDUC	-0.001153	1.92E-06	5.34E-05	-1.56E-06	8.01E-05	1.16E-05
NAGFAM	-0.000607	-4.12E-06	1.94E-05	-9.78E-07	1.16E-05	0.000164

## Equação 2

Dependent Variable: MACBIO

Method: Least Squares

Included observations: 660

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	1.360437	0.155564	8.745168	0.0000
PMACBIO	-0.145798	0.006878	-21.19896	0.0000
PMACNORM	0.155684	0.080862	1.925295	0.0546
FAMREND	0.002611	0.000630	4.144724	0.0000
EDUC	0.026668	0.009359	2.849611	0.0045
NAGFAM	0.082783	0.014569	5.682034	0.0000
FIT^2	0.011123	0.018961	0.586621	0.5577
R-squared	0.469107	Mean dependent var		2.130228
Adjusted R-squared	0.464229	S.D. dependent var		0.676247
S.E. of regression	0.494988	Akaike info criterion		1.441985
Sum squared resid	159.9939	Schwarz criterion		1.489630
F-statistic	96.16701	Durbin-Watson stat		1.900975
Prob(F-statistic)	0.000000			

### Equação 3

Dependent Variable: MACBIO  
Method: Least Squares  
Included observations: 660

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	2.180972	0.056182	38.82006	0.0000
PMACBIO-2.5	-0.146821	0.006650	-22.07964	0.0000
PMACNORM-1.5	0.164934	0.079270	2.080645	0.0379
FAMREND-60	0.002768	0.000570	4.852339	0.0000
EDUC-15	0.028262	0.008951	3.157377	0.0017
NAGFAM-4	0.086866	0.012792	6.790782	0.0000
R-squared	0.468827	Mean dependent var	2.130228	
Adjusted R-squared	0.464766	S.D. dependent var	0.676247	
S.E. of regression	0.494740	Akaike info criterion	1.439481	
Sum squared resid	160.0782	Schwarz criterion	1.480320	
Log likelihood	-469.0288	Hannan-Quinn criter.	1.455311	
F-statistic	115.4474	Durbin-Watson stat	1.903987	
Prob(F-statistic)	0.000000			