

Nome: _____ Nº _____

1. **Perguntas de Verdadeiro/Falso** (2 valores) - Para cada afirmação, assinale se esta é Verdadeira (V) ou Falsa (F). Uma resposta certa vale 0.25 e uma resposta errada penaliza em idêntico valor.

	V	F
Num teste de dimensão $\alpha = 0.1$ em que o valor- $p=0.07$, rejeita-se H_0 .		
$\hat{\theta}$ é estimador não enviesado de θ se e só se $\text{var}(\hat{\theta}) = \text{var}(\theta)$.		
Se num processo repetitivo se forem construindo intervalos de confiança para θ com base em amostras casuais simples independentes umas das outras e todas de mesma dimensão, é legítimo concluir que a proporção de intervalos que vão incluir θ se vai aproximando do nível de confiança escolhido.		
O modelo $z_t = \exp(\alpha_1 + \alpha_2 x_{t2} + \alpha_3 x_{t3} + \alpha_4 x_{t3}^2) \times v_t, t = 1, 2, \dots, n$, é linearizável.		
Num teste do χ^2 à bondade do ajustamento a região de rejeição pode ser bilateral		
Num teste de hipótese simples contra hipótese simples pode acontecer que $P(\text{Rej } H_0 H_0) + P(\text{não Rej } H_0 H_0) < 1$		
A eficiência de um estimador é uma propriedade que apenas interessa em grandes amostras.		
A dimensão de um teste é dada pela probabilidade de rejeitar a hipótese nula quando esta é falsa.		

2. **Perguntas de resposta múltipla** (2 valores) - Para cada pergunta escolha a alternativa correcta. Uma resposta certa vale 0.5 valores e uma resposta errada penaliza em 0.17 valores.

1. A amplitude de um intervalo de confiança para a média de um universo normal de variância conhecida, construído com base no processo habitual, pode ser reduzida por
 - Um aumento da dimensão da amostra
 - Uma diminuição do nível de confiança
 - Qualquer das 2 primeiras alternativas
 - Nenhuma das 2 primeiras alternativas
2. Considere o MRL estimado pelos mínimos quadrados $\hat{Y}_t = b_1 + b_2 x_{t2} + b_3 x_{t3}$, em que Y representa o preço de uma propriedade agrícola em determinada região, x_2 a sua área e x_3 a distância à cidade mais próxima. Qual a estimativa da elasticidade preço-área?
 - b_2
 - $b_2 x_t / \hat{Y}_t$
 - $(b_2 + b_3) x_t / \hat{Y}_t$
 - b_3
3. A Desigualdade de Fréchet-Cramer-Rao é um resultado particularmente importante
 - Na estimação por intervalo
 - Nos testes de hipóteses
 - Na estimação por pontos
 - Para qualquer das alternativas
4. Estimado um MRL pelos mínimos quadrados obteve-se $\ln \hat{y}_t = 5.2 - 0.32 \ln x_{t2} + 3.2 x_{t3}$. Neste quadro quando x_2 aumenta 0.1%, pode-se afirmar que, tudo o resto constante,
 - $E(y | X)$ diminui de 0.032 unidades aproximadamente
 - $E(y | X)$ decresce de 0.32% aproximadamente
 - $E(y | X)$ decresce de 3.2% aproximadamente
 - $E(y | X)$ decresce de 0.032% aproximadamente

3. Perguntas de desenvolvimento (2 valores) – Cada resposta certa vale 1 valor.

- a. Defina o conceito de variável fulcral para o parâmetro θ e diga qual a sua utilidade.
- b. Mostre que, quando se comparam dois modelos com a mesma variável dependente e estimados com base no mesmo conjunto de observações desta, é equivalente escolher aquele que tem maior coeficiente de determinação ajustado ou aquele que tem menor erro padrão da regressão (s^2). Recordar-se que para o modelo $y_t = \beta_1 + \beta_2 x_{t2} + \dots + \beta_k x_{tk} + u_t$, $t = 1, 2, \dots, n$, $\bar{R}^2 = 1 - \frac{VR/(n-k)}{VT/(n-1)}$, sendo VR e VT as variações residuais e totais respectivamente, e $s^2 = VR/(n-k) = \sum_{t=1}^n \hat{u}_t^2 / (n-k)$

Nome: _____ N° _____

Alínea	1a	1b	2	3	4a	4b	4c	4d	4e	4f
Cotação	15	15	15	15	10	10	15	15	15	15
Classificação										

T:

P:

TOTAL:

Em todos os testes de hipóteses que fizer formule as hipóteses, indique a estatística de teste e sua distribuição. Assuma por defeito uma dimensão $\alpha = 0.05$.

1. A distância percorrida por um veículo pesado de mercadorias num dia de trabalho, em centenas de quilómetros, é uma variável aleatória, X , com média $\mu = \theta + 2$, variância $\sigma^2 = \theta^2$ e distribuição dada por

$$f(x|\theta) = \frac{1}{\theta} e^{-\frac{x-2}{\theta}} \text{ para } x > 2, \text{ onde } \theta \text{ é um parâmetro desconhecido } (\theta > 0).$$

- a) Mostre que o estimador de máxima verosimilhança para θ é $\hat{\theta} = \bar{X} - 2$ e analise a sua consistência.
- b) Observada uma amostra casual de dimensão 50 obteve-se $\sum_{i=1}^{50} x_i = 350$. Apresente, justificando, a estimativa da máxima verosimilhança para a percentagem de dias de trabalho em que um veículo desse tipo percorre uma distância superior a 500 quilómetros.

2. Numa empresa de aluguer de veículos o parâmetro fundamental para estabelecer a tarifa diária de aluguer dos veículos ligeiros de passageiros, na categoria de quilometragem ilimitada, é o número médio de quilómetros percorridos diariamente que, na tarifa actualmente praticada, se pressupõe não ultrapassar 275. Para avaliar se é necessário rever essa tarifa recolheu-se uma amostra casual de 500 alugueres nesta categoria, tendo-se observado uma média de 278.5 e uma variância corrigida de 6430.5. Recorra a um teste de hipóteses de dimensão $\alpha = 0.05$ para avaliar se é de proceder à revisão da tarifa diária deste tipo de aluguer.

3. O técnico que está a realizar o estudo defende que existe associação entre a quilometragem média diária percorrida e o número de dias de aluguer. Para testar esta afirmação recorreu a um programa de estatística tendo obtido os seguintes resultados

Dias * Kms Crosstabulation

		Kms					Total
		até 220	220,0 – 250	250,0 – 300	300,0 - 375	mais de 375	
Dias até 7	Count	18	14	55	86	36	209
	Expected Count	35.5	62.7	50.2	? =	? =	209.0
8 ou +	Count	67	136	65	22	1	291
	Expected Count	49.5	87.3	69.8	? =	? =	291.0
Total	Count	85	150	120	108	37	500
	Expected Count	85,0	150.0	120.0	108.0	37.0	500.0

Chi-Square Tests

	Value	Df	Asymp. Sig.
Pearson Chi-Square	191.031 ^a	4	.000
N of Valid Cases	500		

a. 0 cells (0%) have expected count less than 5. The minimum expected count is 15.47.

Complete, no próprio quadro, os valores em falta na tabela de contingência.

Com base no teste de hipóteses adequado diga, apresentando os procedimentos seguidos, se concorda ou não com a afirmação do técnico.

4. A empresa “RentMais” com o propósito de explicar a quilometragem média diária percorrida pelos seus clientes de aluguer de veículos ligeiros de passageiros, construiu o seguinte modelo:

$$KMD_t = \beta_1 + \beta_2 ID_t + \beta_3 ND_t + \beta_4 NAC_t \times ND_t + \beta_5 NAC_t + U_t$$

em que, para o t –ésimo aluguer da amostra, KMD representa a quilometragem média diária, ID a idade do condutor, ND o número de dias contratados e NAC uma variável artificial que assume o valor 1 se o cliente reside em Portugal e 0 caso contrário. Todas as alíneas se referem a este modelo embora possa utilizar os resultados em anexo sempre que necessário.

Em todas as regressões apresentadas o regressando é a quilometragem média diária (KMD).

Supondo satisfeitas as hipóteses do modelo de regressão linear obtiveram-se, utilizando o Excel, os seguintes resultados,

Regression Statistics	
Multiple R	0.7505766
R Square	0.5633653
Adjusted R Square	0.5598369
Standard Error	45.6817893
Observations	500

ANOVA					
	Df	SS	MS	F	Significance F
Regression	4	1332794.623	333198.6559	159.6677	1.19327E-87
Residual	495	1032978.809	2086.8259		
Total	499	2365773.433			

	Coefficients	Standard Error	t Stat	P-value
Intercept	433.1389840	8.5095656	50.9002462	7.7136E-199
ID	-3.4989469	0.2009155	-17.4150212	2.47811E-53
ND	-2.4044767	0.6033652	-3.9851097	7.75984E-05
NDxNAC	-2.9164748	0.7348825	-3.9686275	8.29862E-05
NAC	49.6963674	8.4148690	5.9057803	6.52846E-09

- Analise, com base no valor-p, a relevância estatística individual de cada um dos regressores do modelo. Construa um intervalo de confiança a 95% para β_5 .
- Interprete o valor das estimativas obtidas para β_2 e β_3 .
- Efectue o teste estatístico $H_0 : \beta_4 = \beta_3$ contra $H_1 : \beta_4 \neq \beta_3$. Interprete o seu significado.
- modelo reparametrizado obteve-se $R^2 = 0.620425$ e $\sum \hat{u}_i^2 = 1033902.06$. O que conclui sobre a veracidade da hipótese em questão.
- Calcule o intervalo de previsão a 95% para a quilometragem média diária efectuada por um cliente com 44 anos de idade, não residente em Portugal e que contratou um aluguer para 7 dias.
- Para verificar se a variância da quilometragem média diária é constante estimou-se a regressão auxiliar conducente à realização do teste de White simplificado tendo-se obtido $R^2 = 0.142822$. Explícite qual a regressão referida e diga, formalizando convenientemente, as conclusões a tirar bem como as consequências nas análises feitas anteriormente, se as houver.

ANEXO

<i>Regression Statistics</i>	
Multiple R	0.6589789
R Square	0.4342531
Adjusted R Square	0.4319765
Standard Error	51.8942758
Observations	500

ANOVA					
	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>Significance F</i>
Regression	2	1027344.548	513672.2738	190.742387	3.36205E-62
Residual	497	1338428.885	2693.0159		
Total	499	2365773.433			

	<i>Coefficients</i>	<i>Standard Error</i>	<i>t Stat</i>	<i>P-value</i>
Intercept	440.4983872	8.9268017	49.3456000	1.118E-193
ID	-4.1913316	0.2147825	-19.5143077	2.2055E-63
NAC	3.4037934	4.6712220	0.7286730	0.46654507

<i>Regression Statistics</i>	
Multiple R	0.7504803
R Square	0.5632207
Adjusted R Square	0.5605789
Standard Error	45.6432712
Observations	500

ANOVA					
	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>Significance F</i>
Regression	3	1332452.564	444150.8545	213.194982	8.15341E-89
Residual	496	1033320.869	2083.3082		
Total	499	2365773.433			

	<i>Coefficients</i>	<i>Standard Error</i>	<i>t Stat</i>	<i>P-value</i>
Intercept	434.4453685	7.8674256	55.2207785	5.6368E-214
ID	-3.4848862	0.1977242	-17.6249859	2.41362E-54
AUX1	-2.6323025	0.2175133	-12.1018021	1.01588E-29
NAC	47.1091507	5.4701807	8.6119918	9.60024E-17

Em que $AUX1 = ND + ND \times NAC$

V.S.F.F →

<i>Regression Statistics</i>	
Multiple R	0.7505766
R Square	0.5633653
Adjusted R Square	0.5598369
Standard Error	45.6817893
Observations	500

ANOVA					
	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>Significance F</i>
Regression	4	1332794.623	333198.6559	159.66768	1.19327E-87
Residual	495	1032978.809	2086.8259		
Total	499	2365773.433			

	<i>Coefficients</i>	<i>Standard Error</i>	<i>t Stat</i>	<i>P-value</i>
Intercept	262.3539827	3.1035592	84.5332632	2.5457E-296
ID-44	-3.4989469	0.2009155	-17.4150212	2.47811E-53
ND-7	-2.4044767	0.6033652	-3.9851097	7.75984E-05
NDxNAC	-2.9164748	0.7348825	-3.9686275	8.29862E-05
NAC	49.6963674	8.4148690	5.9057803	6.52846E-09

Nota: A variável ID-44 refere-se, naturalmente, a uma variável obtida subtraindo 44 a cada uma das observações de ID. A definição é semelhante para ND-7.

<i>Regression Statistics</i>	
Multiple R	0.7505766
R Square	0.5633653
Adjusted R Square	0.5598369
Standard Error	45.6817893
Observations	500

ANOVA					
	<i>Df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>Significance F</i>
Regression	4	1332794.623	333198.6559	159.6676845	1.19327E-87
Residual	495	1032978.809	2086.8259		
Total	499	2365773.433			

	<i>Coefficients</i>	<i>Standard Error</i>	<i>t Stat</i>	<i>P-value</i>
Intercept	291.6350264	4.108122813	70.989851	1.1924E-261
ID-44	-3.498946918	0.200915455	-17.41502123	2.47811E-53
ND-7	-2.404476699	0.603365245	-3.985109714	7.75984E-05
NDxNAC -7	-2.916474823	0.734882489	-3.968627454	8.29862E-05
NAC-1	49.69636743	8.414868961	5.905780312	6.52846E-09

Nota: A variável ID-44 refere-se, naturalmente, a uma variável obtida subtraindo 44 a cada uma das observações de ID. A definição é semelhante para ND-7, NDxNAC -7 e NAC-1.