Nome: ______________________________ Nº______________

Espaço reservado a classificações						
1.a)	1.b)	2.a)	2.b)	3.a)	3.b.i)	3.b.ii)
4.a)	4.b)	4.c)	4.d)	4.e)	4.f)	
				TOTAL:		

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!

1. A nova *start-up* Xaquim desenvolveu um modelo de bateria para carros eléctricos que pensa apresentar num projecto à Tesla. O Elon Musk informou que só estaria interessado em tal modelo se tivesse um gasto energético inferior a 5.5 kWh, e então pediu à *start-up* que fosse feito um estudo em que se apresentasse um intervalo de confiança a 99% para a média do gasto energético da bateria. Recolhida uma amostra casual de 50 unidades, obteve-se: $\bar{x} = 5.61$ e $s' = 0.25$.

   a) Calcule e interprete o intervalo de confiança pedido pelo Elon Musk. Acha que o CEO da Tesla estará interessado na bateria da Xaquim? *[2.0]*

IC a 99% para μ (Grandes Amostras: Caso Geral)

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

IC: $\left(\bar{x} \pm z_{0.005} \times s'/\sqrt{n}\right) = \left(5.61 \pm 2.576 \times 0.25/\sqrt{50}\right) = (5.519, 5.701)$

Com uma confiança de 99%, podemos concluir que a média do gasto energético da bateria da Xaquim se encontra entre 5.519 kWh e 5.701 kWh. Porque o IC obtido contém valores superiores a 5.5 kWh, conclui-se que o CEO da Tesla não estará interessado nesta bateria da Xaquim.

   b) De acordo com o intervalo de confiança anterior, o grau de confiança de 99% significa que…
   [Resposta certa: 1 / Resposta errada: -0.25]

… a amplitude do intervalo de confiança é de 0.99.	
… a probabilidade do intervalo de confiança conter a verdadeira média é de 0.99.	
… a probabilidade da verdadeira média estar dentro do intervalo aleatório é de 0.99.	**X**
… em 100 intervalos construídos, 99 são de confiança.	

2. A Xaquim terminou um projecto de outro modelo de bateria que também pretende apresentar à Tesla, mas só o decidirá fazer se mais de 70% dos seus clientes estiverem satisfeitos. Tendo recolhido uma amostra casual de 100 clientes, 78 responderam estar satisfeitos enquanto os restantes afirmaram não estar satisfeitos.

a) Qual será a conclusão da Xaquim, considerando um nível de 5%? *[2.0]*

<u>População de Bernoulli</u>

$$H_0: \theta \leq 0.7 \quad \text{vs} \quad H_1: \theta > 0.7$$

$$\textbf{ET}: Z = \frac{\bar{X} - \theta_0}{\sqrt{\frac{\theta_0(1-\theta_0)}{n}}} \overset{a}{\sim} N(0,1)$$

$$W_{0.05} = \{z_{obs}: z_{obs} > 1.645\}$$

$$z_{obs} = \frac{\left(\frac{78}{100}\right) - 0.7}{\sqrt{\frac{0.7(1-0.7)}{100}}} \approx 1.746$$

Porque $z_{obs} \in W_{0.05}$, rejeita-se H_0 ao nível de 5%. Evidência estatística favorável à hipótese de mais de 70% dos clientes da Xaquim estarem satisfeitos com esta bateria. Logo, a Xaquim deverá apresentar este modelo à Tesla.

b) A dimensão de 5% do teste anterior significa que…
[Resposta certa: 1 / Resposta errada: -0.25]

… em 100 amostras casuais, o teste rejeita erradamente H_0 em 5.	**X**
… a probabilidade de não rejeitar H_0 quando H_0 é verdadeira é de 0.05.	
… o respectivo valor-p é de 0.95.	
… a probabilidade de a dimensão da amostra ser grande é de 5%.	

3. Assuma que o gasto energético de uma bateria é uma variável aleatória X positiva com distribuição desconhecida de média $\mu > 0$ e variância σ^2. O Elon Musk enviou um *tweet* à Xaquim em que sugeria o seguinte estimador para a média do gasto energético:

$$\frac{1}{n}\sum_{i=1}^{n} \ln\left(\frac{X_i}{1+X_i}\right)$$

a) Usando este estimador, em que intervalo de valores se encontra qualquer estimativa para μ? Diga, justificando, se faz ou não sentido usar este estimador para μ. *[1.0]*

Sabe-se que a variável aleatória X é positiva, pelo que:

$$0 < x_i < 1 + x_i \Rightarrow 0 < \frac{x_i}{1+x_i} < 1 \Rightarrow \ln\left(\frac{x_i}{1+x_i}\right) < 0 \Rightarrow \sum_{i=1}^{n} \ln\left(\frac{x_i}{1+x_i}\right) < 0 \Rightarrow \frac{1}{n}\sum_{i=1}^{n} \ln\left(\frac{x_i}{1+x_i}\right) < 0$$

Isto significa que usando este estimador qualquer estimativa para μ estará no intervalo $]-\infty, 0[$, isto é, qualquer estimativa para μ será sempre negativa, pelo que não faz sentido usar este estimador uma vez que $\mu > 0$.

b) Assuma que a Xaquim recolheu uma amostra casual de baterias de dimensão n para obter uma estimativa pontual de μ, mas ao analisar os dados, o novo estagiário perdeu inadvertidamente o valor do gasto energético da última observação. Neste cenário, a Xaquim irá optar por um dos seguintes estimadores:

$$T_1 = \frac{1}{n-1}\sum_{i=1}^{n-1} X_i \quad \text{e} \quad T_2 = \frac{1}{n}\sum_{i=1}^{n-1} X_i$$

Em relação ao estimador T_1, sabe-se que é centrado e consistente para μ.

i) Mostre que o estimador T_2 é enviesado, mas consistente para μ. *[2.0]*

<u>Enviesamento</u>:

$$E(T_2) = E\left(\frac{1}{n}\sum_{i=1}^{n-1} X_i\right) = \frac{1}{n}E\left(\sum_{i=1}^{n-1} X_i\right) = \frac{1}{n}\sum_{i=1}^{n-1} E(X_i) = \frac{1}{n}\sum_{i=1}^{n-1}\mu = \frac{1}{n}(n-1)\mu = \mu - \frac{\mu}{n} \neq \mu$$

Logo, o estimador T_2 é enviesado para μ.

<u>Consistência</u>:

$$\lim_{n\to\infty} E(T_2) = \lim_{n\to\infty}\left(\mu - \frac{\mu}{n}\right) = \mu - 0 = \mu$$

Note-se que:

$$Var(T_2) = Var\left(\frac{1}{n}\sum_{i=1}^{n-1} X_i\right) = \left(\frac{1}{n}\right)^2 Var\left(\sum_{i=1}^{n-1} X_i\right) = \frac{1}{n^2}\sum_{i=1}^{n-1} Var(X_i) = \frac{1}{n^2}\sum_{i=1}^{n-1}\sigma^2 = \frac{1}{n^2}(n-1)\sigma^2 =$$
$$= \frac{\sigma^2}{n} - \frac{\sigma^2}{n^2}$$

Logo,

$$\lim_{n\to\infty} Var(T_2) = \lim_{n\to\infty}\left(\frac{\sigma^2}{n} - \frac{\sigma^2}{n^2}\right) = 0$$

Porque $\lim_{n\to\infty} E(T_2) = \mu$ e $\lim_{n\to\infty} Var(T_2) = 0$, conclui-se que o estimador T_2 é consistente para μ.

ii) Sabe-se que $Var(T_2) < Var(T_1)$. Então…
[Resposta certa: 1 / Resposta errada: -0.25]

… o estimador T_2 é mais eficiente que o estimador T_1.	
… nada se pode concluir quanto à eficiência.	**X**
… o estimador T_1 é mais eficiente que o estimador T_2.	

4. No seu papel de consultor estatístico, admita que foi contratado por um cliente para estudar o comportamento das vendas de um determinado produto em vários mercados com base nos gastos em publicidade. O modelo adoptado é o seguinte:

$$\ln(VENDAS) = \beta_1 + \beta_2 \ln(TV) + \beta_3 \ln(RADIO) + \beta_4 \ln(JORNAL) + u$$

Onde:

- $\ln(VENDAS)$ – logaritmo da quantidade vendida (milhares de unidades) do produto;
- $\ln(TV)$ – logaritmo do montante (milhares de euros) gasto em publicidade na televisão;
- $\ln(RADIO)$ – logaritmo do montante (milhares de euros) gasto em publicidade na rádio;
- $\ln(JORNAL)$ – logaritmo do montante (milhares de euros) gasto em publicidade nos jornais.

Todas as questões que se seguem referem-se sempre ao modelo inicial cuja estimação pode encontrar no **ANEXO** na **Equação 1**. No **ANEXO** também se encontram estimações de outras regressões que deve utilizar quando necessário como auxiliares de cálculo.

a) Interprete as estimativas dos parâmetros associados aos montantes gastos em publicidade na televisão e no jornal, e teste a significância individual de ambas as variáveis ao nível de 5%. *[2.0]*

Interpretação:

$b_2 = 0.34935 \rightarrow$ *Ceteris paribus*, estima-se em média que se o montante gasto em publicidade na televisão aumentar 1%, a quantidade vendida do produto aumenta aproximadamente 0.34935%.

$b_4 = 0.01594 \rightarrow$ *Ceteris paribus*, estima-se em média que se o montante gasto em publicidade nos jornais aumentar 1%, a quantidade vendida do produto aumenta aproximadamente 0.01594%.

Significância individual de ln(*TV*):

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

$$\textbf{ET}: t_2 = \frac{b_2}{s_{b_2}} \sim t(196)$$

Do output obtém-se $p_{2,obs} = 0$. Porque $p_{2,obs} < \alpha = 0.05$, rejeita-se H_0 ao nível de 5%. Logo, o regressor ln(*TV*) é estatisticamente significativo a 5%.

Significância individual de ln(*JORNAL*):

$$H_0: \beta_4 = 0 \quad \text{vs} \quad H_1: \beta_4 \neq 0$$

$$\textbf{ET}: t_4 = \frac{b_4}{s_{b_4}} \sim t(196)$$

Do output obtém-se $p_{4,obs} = 0.06495$. Porque $p_{4,obs} > \alpha = 0.05$, não se rejeita H_0 ao nível de 5%. Logo, o regressor ln(*JORNAL*) não é estatisticamente significativo a 5%.

b) Construa um intervalo de confiança a 95% para o parâmetro β_3 e interprete. *[1.0]*

IC a 95% para β_3:

$$\mathbf{VF}: t_3 = \frac{b_3 - \beta_3}{s_{b_3}} \sim t(196)$$

IC: $\left(b_3 \pm z_{0.025} \times s_{b_3}\right) = (0.15896 \pm 1.96 \times 0.00767) = (0.144, 0.174)$

Com uma confiança de 95%, podemos concluir que a elasticidade da quantidade vendida do produto em relação ao montante gasto em publicidade na rádio se situa entre 0.144% e 0.174%.

c) Avalie a significância global da **Equação 1** ao nível de 5%. O que pode concluir? *[1.0]*

Significância Global:

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

$$\mathbf{ET}: F = \frac{R^2/3}{(1 - R^2)/196} \sim F(3, 196)$$

Do output obtém-se: $p_{obs} = 0 < \alpha = 0.05$, pelo que se rejeita H_0 ao nível de 5%. Logo, conclui-se que o modelo é globalmente significativo.

d) Teste a 5% a seguinte afirmação: "tudo o resto constante, o efeito sobre as vendas de um aumento de 1% nos gastos de publicidade na televisão é maior que o dobro do efeito de um aumento idêntico nos gastos em publicidade na rádio." O que pode concluir quanto à veracidade desta afirmação? (**<u>Nota</u>**: Utilize a **Equação 2** como auxiliar de cálculos.) *[2.0]*

$$H_0: \beta_2 \leq 2\beta_3 \quad \text{vs} \quad H_1: \beta_2 > 2\beta_3$$

Devemos reparametrizar o modelo inicial. Sob H_0, $\beta_2 \leq 2\beta_3$, que é equivalente a: $\beta_2 - 2\beta_3 \leq 0$. Seja então $\delta = \beta_2 - 2\beta_3$. Logo,

$$H_0: \delta \leq 0 \quad \text{vs} \quad H_1: \delta > 0$$

$$\textbf{ET}: t_{\hat{\delta}} = \frac{\hat{\delta}}{s_{\hat{\delta}}} \sim t(196)$$

$$W_{0.05} = \{t_{\hat{\delta},obs}: t_{\hat{\delta},obs} > 1.645\}$$

Para obter o modelo reparametrizado note-se que: $\delta = \beta_2 - 2\beta_3 \Leftrightarrow \beta_2 = \delta + 2\beta_3$. Então, substituindo no modelo inicial, obtém-se o modelo reparametrizado:

$$\ln(VENDAS) = \beta_1 + (\delta + 2\beta_3)\ln(TV) + \beta_3 \ln(RADIO) + \beta_4 \ln(JORNAL) + u \Leftrightarrow$$
$$\Leftrightarrow \ln(VENDAS) = \beta_1 + \delta \ln(TV) + \beta_3[\ln(RADIO) + 2\ln(TV)] + \beta_4 \ln(JORNAL) + u$$

que corresponde à Equação 2 em Anexo, e da qual se obtém directamente: $t_{\hat{\delta},obs} = 1.79488$. Logo, porque $t_{\hat{\delta},obs} \in W_{0.05}$, rejeita-se H_0 ao nível de 5%, existindo evidência estatística favorável à afirmação.

e) Construa um intervalo de previsão a 99% para a quantidade vendida do produto num determinado mercado com os seguintes montantes gastos em publicidade: 150 mil euros em televisão, 25 mil euros em rádio e 30 mil euros em jornal. Interprete o resultado obtido.
(**<u>Nota</u>**: Utilize a **Equação 3** como auxiliar de cálculos.) *[2.0]*

<u>Previsão Pontual</u>:

IP a 99% para $\ln(VENDAS)$:

$$\textbf{VF}: \ t_d = \frac{\ln(VENDAS_0) - \ln(\widehat{VENDAS}_0)}{s_d} \sim t(196)$$

Usando a Equação 3 em Anexo, obtém-se:

$$\left(\ln(\widehat{VENDAS}_0) \pm t_{0.005} \times s_d\right) = \left(2.75816 \pm 2.576 \times \sqrt{0.11569^2 + 0.00946^2}\right) = (2.45915, 3.05717)$$

IP a 99% para $VENDAS$:

$$(e^{2.45915}, e^{3.05717}) = (11.695, 21.267)$$

Logo, com uma confiança de 99% prevê-se que a quantidade vendida do produto num mercado com estes montantes de publicidade se situa entre 11.695 milhares de unidades e 21.267 milhares de unidades.

f) Para estudar a possível existência de comportamentos diferenciados na quantidade vendida do produto consoante a dimensão dos mercados, estimou-se separadamente dois modelos idênticos ao inicial: um com observações de grandes mercados do qual resultou $VR = 0.10861$; e outro com observações de pequenos mercados em que se obteve $VR = 1.05562$. Efectue um teste de hipóteses adequado ao nível de 5% e conclua. *[2.0]*

<u>Teste de Chow</u>:

$$H_0: \forall j, \beta_{jG} = \beta_{jP}\ (j = 1, 2, 3, 4) \quad \text{vs} \quad H_1: \exists j: \beta_{jG} \neq \beta_{jP}\ (j = 1, 2, 3, 4)$$

$$\textbf{ET}: F_{Chow} = \frac{(VR_0 - VR_1)/k}{VR_1/(n-2k)} \sim F(k, n-2k) = F(4, 192)$$

$$W_{0.05} = \{F_{obs}: F_{obs} > 2.37\}$$

$$F_{obs} = \frac{[2.62314 - (0.10861 + 1.05562)]/4}{(0.10861 + 1.05562)/192} \approx 60.149$$

Porque $F_{obs} \in W_{0.05}$, rejeita-se H_0 ao nível de 5%. Evidência estatística favorável à alteração da estrutura, o que sugere a existência de comportamentos diferenciados na quantidade vendida do produto consoante a dimensão dos mercados.

ANEXO

Equação 1:

$$\ln(VENDAS) = \beta_1 + \beta_2 \ln(TV) + \beta_3 \ln(RADIO) + \beta_4 \ln(JORNAL) + u$$

Regression Statistics	
Multiple R	0.96084
R Square	0.92322
Adjusted R Square	0.92205
Standard Error	0.11569
Observations	200

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regressão	3	31.54247	10.51416	785.61330	0.00000
Residual	196	2.62314	0.01338		
Total	199	34.16562			

	Coefficients	*Standard Error*	*t Stat*	*P-value*
Intercept	0.44182	0.04838	9.13195	0.00000
$\ln(TV)$	0.34935	0.00817	42.76287	0.00000
$\ln(RADIO)$	0.15896	0.00767	20.72162	0.00000
$\ln(JORNAL)$	0.01594	0.00859	1.85603	0.06495

Equação 2:

Regressando: $\ln(VENDAS)$

Regression Statistics	
Multiple R	0.96084
R Square	0.92322
Adjusted R Square	0.92205
Standard Error	0.11569
Observations	200

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	3	31.54247	10.51416	785.61330	0.00000
Residual	196	2.62314	0.01338		
Total	199	34.16562			

	Coefficients	*Standard Error*	*t Stat*	*P-value*
Intercept	0.44182	0.04838	9.13195	0.00000
$\ln(TV)$	0.03144	0.01752	1.79488	0.07421
$\ln(RADIO) + 2\ln(TV)$	0.15896	0.00767	20.72162	0.00000
$\ln(JORNAL)$	0.01594	0.00859	1.85603	0.06495

Equação 3:

Regressando: $\ln(VENDAS)$

Regression Statistics	
Multiple R	0.96084
R Square	0.92322
Adjusted R Square	0.92205
Standard Error	0.11569
Observations	200

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	3	31.54247	10.51416	785.61330	0.00000
Residual	196	2.62314	0.01338		
Total	199	34.16562			

	Coefficients	*Standard Error*	*t Stat*	*P-value*
Intercept	2.75816	0.00946	291.66454	0.00000
ln(TV) – ln(150)	0.34935	0.00817	42.76287	0.00000
ln(RADIO) – ln(25)	0.15896	0.00767	20.72162	0.00000
ln(JORNAL) – ln(30)	0.01594	0.00859	1.85603	0.06495