Table 1: **RESOLUÇÃO 1** (2 points cada resposta exacta, -1 errada)

		V	F
1	Um intervalo de confiança a 95% para θ não contém o verdadeiro valor de θ, com probabilidade 0.05.	X	
2	Num modelo de regressão linear múltipla, um valor do $R^2 = 0.5$ significa que a proporção da variação da variável dependente explicada pela variação dos regressores do modelo é de 50%.	X	
3	Considere o modelo $\log(y_i) = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + u_i$. Neste caso, o efeito parcial aproximado de x sobre y é dado por $\%\Delta y = (\beta_1 + \beta_2)\Delta x \times 100\%$.		X
4	Quando se acrescentam variáveis explicativas ao modelo, o R^2 so pode decrescer.		X
5	Considere o modelo $y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + u_i$, em que a correlação entre x_1 e x_2 é igual a 0.999. Nesse caso, o modelo sofre de multicolinearidade.	X	
6	Um estimador $\hat{\theta}$ para um parametro θ é negativamente enviesado se $E[\hat{\theta}] < \theta$.	X	
7	O valor-p, tambem chamado nível de significância associado à estatística teste observada, é a probabilidade de obter um valor tão ou mais extremo/desfavorável para a hipótese alternativa do que o observado para uma amostra obtida.		X
8	Sob as hipóteses do teorema de Gauss-Markov, não é possível encontrar outro estimador com variância mais baixa que o estimador OLS.	X	
9	É possível utilizar directamente o método OLS para estimar o modelo $y_i = \exp(\beta_0 + \beta_1 x_i)\exp(u_i)$.		X
10	O erro de 2ª espécie consiste em rejeitar H_0, quando H_0 é verdadeira.		X

RESOLUÇÃO 2.a)

$$\begin{cases} E(X) = \overline{X} \\ E(X^2) = \sum_{i=1}^n X_i^2/n \end{cases} \Leftrightarrow \begin{cases} \alpha = \overline{X} \\ 2\beta + \alpha^2 = \sum_{i=1}^n X_i^2/n \end{cases} \Rightarrow \begin{cases} \tilde{\alpha} = \overline{X} \\ \tilde{\beta} = (\sum_{i=1}^n X_i^2/n - \tilde{\alpha}^2)/2 \end{cases}$$

$$\Leftrightarrow \begin{cases} \tilde{\alpha} = \overline{X} \\ \tilde{\beta} = (\sum_{i=1}^n X_i^2/n - \overline{X}^2)/2 = S^2/2 \end{cases}$$

RESOLUÇÃO 2.b)

$E(\tilde{\alpha}) = E(\overline{X}) = E(X) = \alpha \to \tilde{\alpha}$ é um estimador centrado para α.

$E(\tilde{\beta}) = E\left(\frac{S^2}{2}\right) = \frac{1}{2}E(S^2) = \frac{1}{2}\frac{n-1}{n} \cdot 2\beta = \frac{n-1}{n}\beta \to \tilde{\beta}$ é um estimador enviesado para β.

Seja

$$T = \frac{n}{n-1}\tilde{\beta} = \frac{n}{2(n-1)}S^2$$

Logo, $E(T) = \beta$, pelo que T é um estimador centrado para β.

RESOLUÇÃO 2.c)

$$\textbf{VF: } Z = \frac{\overline{X} - \alpha}{S/\sqrt{100}} \overset{a}{\sim} N(0,1) \longrightarrow \text{IC a 99\% para } \alpha: \left(\overline{x} \mp 2.576\frac{s}{\sqrt{100}}\right)$$

Sabendo que $\overline{x} = 16.65/100 = 0.1665$ e que $s = \sqrt{70.35/100 - 0.1665^2} \approx 0.822$, obtém-se o intervalo: $(-0.045, 0.378)$.

RESOLUÇÃO 3.

$$H_0 : \sigma = 0.025 \ \text{ vs } \ H_1 : \sigma > 0.025$$

$$\textbf{ET: } Q = \frac{29 \cdot S'^2}{0.025^2} = 46400 S'^2 \sim \chi(29)$$

$$W_{0.05} = \{q_{obs} : q_{obs} > 42.557\}$$

Como $q_{obs} = 46400 \times 0.031^2 = 44.5904 \in W_{0.05}$, rejeita-se H_0 ao nível de 5%. Evidência estatística de que a máquina necessita ser ajustada.

RESOLUÇÃO 4.
Trata-se de um teste nao-parametrico de independencia

$$H_0 : \forall(i,j) p_{ij} = p_{i\circ} p_{\circ j}, \ \text{ vs } \ H_1 : H_0 \text{ falsa}$$

$$Q = \sum_{i=1}^{2}\sum_{j=1}^{2} \frac{(N_{ij} - n\hat{p_{i\circ}}\hat{p_{\circ j}})^2}{n\hat{p_{i\circ}}\hat{p_{\circ j}}} \sim \chi^2(1)$$

$\hat{p}_{sensivelA} = 0.56 \quad \hat{p}_{resistenteA} = 0.44 \quad \hat{p}_{sensivelB} = 0.36 \quad \hat{p}_{resistenteB} = 0.64$

$$Q_{obs} = \frac{(52-42.34)^2}{42.34} + \frac{(23-33.26)^2}{33.26} + \frac{(65-75.26)^2}{75.26} + \frac{(70-59.14)^2}{59.14} = 2.20 + 3.16 + 1.40 + 1.99 = 8.76$$

Regiao de rejeicao para teste unilateral $W = \{Q : Q > 3.841\}$
Valor-p$=P(Q > 8.76) = 0.99$

Rejeita-se H_0, ao nível de 5%: as duas resistencias antimicrobianas nao sao independentes.

RESOLUÇÃO 5.

Considere o seguinte modelo para estudar a taxa de criminalidade num conjunto de freguesias das Áreas Metropolitanas de Lisboa e do Porto, no ano de 2024:

$$txcrim_i = \beta_0 + \beta_1 \log(rend_i) + \beta_2 \log(pmc_i) + \beta_3\, prppob_i + u_i \tag{1}$$

onde *txcrim* corresponde à taxa de criminalidade (dividida por 100) na freguesia i, *rend* corresponde ao rendimento mediano anual (em euros) na freguesia i, *pmc* corresponde ao preço mediano (em euros) das casas na freguesia i e *prppob* corresponde à proporção de habitantes (dividida por 100) que vivem no (ou abaixo do) limiar da pobreza na freguesia i. Notar que a função $\log(\cdot)$ corresponde ao logaritmo natural (base e). Assuma que o erro, u, tem média nula.

Com base numa amostra aleatória de dimensão 136 e com recurso ao método dos Mínimos Quadrados Ordinários (OLS), foram obtidas as seguintes estimativas (com os erros-padrão dentro de parêntesis):

$$\begin{aligned} \widehat{txcrim}_i &= \underset{(0.192)}{-0.876} - \underset{(0.012)}{0.004}\log(rend_i) - \underset{(0.010)}{0.032}\log(pmc_i) + \underset{(0.044)}{0.765}\,prppob_i \\ R^2 &= 0.696, \quad SSR = 0.279 \end{aligned} \tag{2}$$

(a) (10 points) Interprete a estimativa $\hat{\beta}_1$ e teste a sua significância estatística considerando $\alpha = 0.1$.

Resposta:

Um acréscimo de 1% no rendimento mediano anual induz, em média, ceteris paribus, uma diminuição de 0.004 pontos percentuais na taxa de criminalidade (note que $txcrim$ está dividida por 100; ver matéria referente à alteração de escala na variável dependente) .

$$H_0 : \beta_1 = 0 \quad \text{contra} \quad H_1 : \beta_1 \neq 0$$

$$T = \frac{\hat{\beta}_1 - \beta_1}{\hat{\sigma}_{\widehat{\beta}_1}} = \frac{\hat{\beta}_1}{\hat{\sigma}_{\widehat{\beta}_1}} \sim t_{(136-3-1)} \xrightarrow{d} N(0,1)$$

$$W_{0.05} = \{t : t < -1.645 \text{ ou } t > 1.645\}$$

$$t_{obs} = \frac{-0.004}{0.012} \approxeq -0.333$$

logo não se rejeita a hipotése nula ao nível de significância de 5%. A variável $rend$ não tem relevância estatística .

(b) (20 points) Com base no modelo estimado em 2 , fizeram-se as seguintes afirmações:

- em média, o aumento do preço mediano das casas (pmc) faz diminuir a taxa de criminalidade, *ceteris paribus*;
- em média, a diminuição da proporção de habitantes que vivem no limiar da pobreza ($prppob$) faz diminuir a taxa de criminalidade, *ceteris paribus*.

Apresente a estatística de teste que permite testar individualmente cada uma das afirmações (identifique as respetivas hipóteses), efetue os testes estatísticos ($\alpha = 0.05$) e mostre as suas conclusões.

Resposta:

1ª afirmação (efeito de *pmc*):

$$H_0 : \beta_2 \geq 0 \quad \text{contra} \quad H_1 : \beta_2 < 0$$

$$T = \frac{\widehat{\beta}_2 - \beta_2}{\widehat{\sigma}_{\widehat{\beta}_2}} = \frac{\widehat{\beta}_2}{\widehat{\sigma}_{\widehat{\beta}_2}} \sim t(132) \xrightarrow{d} N(0,1)$$

$$W_{0.05} = \{t : t < -1.645\}$$

$$t_{obs} = \frac{-0.032}{0.010} \approxeq -3.2$$

logo rejeita-se claramente a hipótese nula e pode concluir-se que um aumento do preço mediano origina, em média uma diminuição na taxa de criminalidade.

2ª afirmação (efeito de *prppob*):

$$H_0 : \beta_3 \leq 0 \quad \text{contra} \quad H_1 : \beta_3 > 0$$

$$T = \frac{\widehat{\beta}_3 - \beta_3}{\widehat{\sigma}_{\widehat{\beta}_3}} = \frac{\widehat{\beta}_3}{\widehat{\sigma}_{\widehat{\beta}_3}} \sim t(132) \xrightarrow{d} N(0,1)$$

$$W_{0.05} = \{t : t > 1.645\}$$

$$t_{obs} = \frac{0.765}{0.044} \approxeq 17.386$$

logo rejeita-se claramente a hipótese nula e pode concluir-se que um aumento da proporção de habitantes que vivrm no limiar da pobreza origina, em média um aumento na taxa de criminalidade.

(c) (10 points) Construa o intervalo de confiança a 90% para β_2. O que pode concluir sobre a significância estatística deste coeficiente?

Resposta:

O intervalo de confiança com 90% de confiança para β_2 é dado por:

$$\left(\widehat{\beta}_2 - z_{0.05} \times \widehat{\sigma}_{\beta_2} \, ; \, \widehat{\beta}_2 + z_{0.05} \times \widehat{\sigma}_{\beta_2} \right) = (\, -0.032 - 1.645 \times \ 0.010 \, ; \, -0.032 - 1.645 \times \ 0.010 \,)$$

$$= (\, -0.04845 ; -0.0156 \,)$$

Como o intervalo de confiança não inclui o zero, o coeficiente é estatisticamente significativo ao nível de 10% .

(d) (15 points) Explique como testaria a $H_0 : \beta_1 = -\beta_2 - 0.04$ contra $H_1 : \beta_1 > -\beta_2 - 0.04$ no modelo 2 , reparametrizando o modelo definido acima. Identifique a estatística de teste e a respetiva distribuição.

Resposta: Defina $\delta = \beta_1 + \beta_2$, o que significa que $\beta_1 = \delta - \beta_2$.

Logo,

$$\begin{aligned} txcrim_i &= \beta_0 + \beta_1 \log(rend_i) + \beta_2 \log(pmc_i) + \beta_3\, prppob_i + u_i \\ &= \beta_0 + (\delta - \beta_2) \log(rend_i) + \beta_2 \log(pmc_i) + \beta_3\, prppob_i + u_i \\ &= \beta_0 + \delta \log(rend_i) + \beta_2 \big(\log(pmc_i) - \log(rend_i) \big) + \beta_3\, prppob_i + u_i \end{aligned}$$

Fazer inferência sobre δ a partir do output obtido para a regressão acima: regressão de $txcrim$ sobre $\log(rend)$, $\log(pmc) - \log(rend)$ e $prppob$.

$$T = \frac{\widehat{\delta} + 0.04}{\widehat{\sigma}_\delta} \sim t(132) \xrightarrow{d} N(0,1)$$

(e) (20 points) Considere a seguinte equação estimada com recurso ao OLS:

$$\begin{aligned} \widehat{txcrim_i + 0.1 \log}(rend_i) &= \underset{(0.010)}{-1.214} + \underset{(0.045)}{0.746} \log(pmc_i) \\ R^2 = 0.675&, \quad SSR = 0.300 \end{aligned} \tag{3}$$

Esta equação visa testar uma hipótese sob dois parâmetros. Identifique-a. Apresente a estatística de teste e a respetiva distribuição. Apresente ainda a conclusão estatística, assumindo o nível de significância $\alpha = 0.05$.

Resposta:

$H_0 : \beta_1 = -0.1$ e $\beta_3 = 0$

A equação 3 dá-nos as estimativas para o modelo com restrições (sob H_0) e a equação 2 dá-nos as estimativas para o modelo com restrições. Portanto, tem-se que $SSR_* = 0.300$ e $SSR = 0.279$.

$$F = \frac{(SSR - SSR_*)/2}{SSR/132} \sim F(2, 132)$$

$$W_{0.05} = \{f : f > 3.07\}$$

$$f = \frac{0.300 - 0.279/2}{0.279/132} = \frac{0.0105}{0.00211} \approx 4.98$$

Como $f \in W_{0.05}$ (ou valor-p < 0.05, para quem tenha optado por esta via), rejeita-se a hipótese nula H_0 .