ISEG – Lisbon School of Economics and Management
ESTATÍSTICA II
2º Semestre 2024/2025
Exame – Época Normal
07 de Maio de 2025

Duração: **2 horas**

Nome:

Questão	**1**	**2**	**3**	**4**	**5**	Total
Pontuação	20	40	25	40	75	200

- **Justifique** todas as suas respostas. Apresente todas as fórmulas utilizadas, e os respetivos cálculos auxiliares;
- Entregue as suas respostas nas **folhas de teste**, escrevendo o seu **nome e número** de aluno em todas;
- **Organize as suas respostas**. Respostas espalhadas e rasuradas na folha de teste, receberão pouco crédito;
- **É permitida a utilização** do formulário da disciplina (sem qualquer anotação), das tabelas estatísticas e de uma calculadora científica. Considere sempre um **nível de significância de 5**%, a não ser que outro nível de significância esteja especificado.
- **Verdadeiro ou Falso:** resposta certa vale 2 pontos; resposta errada vale −1 ponto; ausência de resposta ou resposta dúbia vale 0 pontos.

VERDADEIRO OU FALSO

(20) **1**. Das afirmações seguintes, assinale com uma cruz (**X**) as que são verdadeiras e falsas.

		V	**F**
1	Um intervalo de confiança a 95% para θ não contém o verdadeiro valor de θ, com probabilidade 0.05.		
2	Num modelo de regressão linear múltipla, um valor do $R^2 = 0.5$ significa que a percentagem da variação da variável dependente explicada pela variação dos regressores do modelo é de 50%.		
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 $\%\Delta y = (\beta_1 + \beta_2)\,\Delta x \times 100\%$.		
4	Quando se acrescentam variáveis explicativas ao modelo, o R^2 só pode decrescer.		
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.		
6	Um estimador $\hat{\theta}$ para um parametro θ é negativamente enviesado se $E[\hat{\theta}] < \theta$.		
7	O valor-p, também 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.		
8	Sob as hipóteses do teorema de Gauss-Markov, não é possível encontrar outro estimador com variância mais baixa do que o estimador OLS.		
9	É possível utilizar diretamente o método OLS para estimar o modelo $y_i = \exp(\beta_0 + \beta_1 x_i)\exp(u_i)$.		
10	O erro de 2ª espécie consiste em rejeitar H_0, quando H_0 é verdadeira.		

RESPOSTA ABERTA

2. De uma população X cuja distribuição depende dos parâmetros $\alpha \in \mathbb{R}$ e $\beta > 0$, sabe-se que $E(X) = \alpha$ e $Var(X) = 2\beta$. Seja $(X_1, \ldots, X_n)$ uma amostra casual desta população.

(10) (a) Mostre que os estimadores do método dos momentos para α e β são, respetivamente, $\overline{X}$ e $S^2/2$.

(15) (b) Analise o enviesamento de ambos os estimadores. O que pode concluir? Se necessário, corrija o enviesamento propondo um novo estimador que seja centrado.

(15) (c) Sabendo que de uma amostra particular de dimensão 100 se obteve $\sum_{i=1}^{100} x_i = 16.65$ e $\sum_{i=1}^{100} x_i^2 = 70.35$, use a distribuição por amostragem assintótica da média para construir um intervalo de confiança a 99% para α.

(25) **3**. As folhas de papel para cadernos produzidas por determinada fábrica provêm de uma máquina que é monitorizada periodicamente devido a possíveis flutuações na grossura das folhas. Assume-se que a grossura de uma folha de papel produzida por esta máquina segue uma distribuição normal. Se o verdadeiro desvio padrão exceder 0.025 milímetros, considera-se que a máquina não está afinada e, como tal, deverá ser reparada. Sabendo que num caderno com 30 folhas se obteve $s' = 0.031$, efetue um teste estatístico adequado que lhe permita averiguar se a máquina em questão deve ou não ser reparada. Considere um nível de significância de 5%.

(40) **4**. Com o objetivo de estudar a relação entre a resistência bacteriana a um antibiótico A e a resistência a um antibiótico B, foram selecionados de forma aleatória 210 pacientes com infeções, tendo-se obtido as seguintes frequências absolutas:

Número de pacientes		**Antibiótico A**	
		Sensível	**Resistente**
Antibiótico B	**Sensível**	52	23
	Resistente	65	70

Com base num teste estatístico de dimensão 0.05, pode-se afirmar que as duas resistências estão associadas?

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:

$$txcrim = \beta_0 + \beta_1 \ln(rend) + \beta_2 \ln(pmc) + \beta_3\, prppob + u \tag{1}$$

onde, para cada freguesia, *txcrim* corresponde à taxa de criminalidade (dividida por 100), *rend* corresponde ao rendimento mediano anual (em euros), *pmc* corresponde ao preço mediano (em euros) das casas e *prppob* corresponde à proporção de habitantes (dividida por 100) que vivem no (ou abaixo do) limiar da pobreza. 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):

$$\begin{aligned} \widehat{txcrim} &= \underset{(0.192)}{-0.876} - \underset{(0.012)}{0.004} \ln(rend) - \underset{(0.010)}{0.032} \ln(pmc) + \underset{(0.044)}{0.765}\, prppob \\ R^2 &= 0.696, \quad SSR = 0.279 \end{aligned} \tag{2}$$

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

(20) (b) 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.

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

(15) (d) Explique como testaria $H_0 : \beta_1 = -\beta_2 - 0.04$ 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.

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

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

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