

LICENCIATURA EM CIÊNCIA DE DADOS

Análise Exploratória de Dados

Teste/Exame de 1ª época

02 de junho 2020

Duração: 2h

Nome:

Número de aluno:

Turma:

Exercício 1 (1,5 valores)

Na tabela abaixo está representada a evolução da população residente em Portugal entre 1978 e 2018 por grupos etários:

Grupos etários	Anos				
	1978	1988	1998	2008	2018
0-14	2559245	2192503	1707310	1637410	1415732
15-44	3925280	4351788	4492201	4331043	3726609
45-64	421830	524145	555796	590099	663927
65 ou +	1036005	1269885	1597033	1884333	2228751
Total	9558250	10019610	10160196	10558177	10283822

Fontes: INE - Estimativas Anuais da População Residente; Fonte: Pordata

(0,75) a) Verifique, justificando com cálculos, se a afirmação seguinte é verdadeira ou falsa:

" Entre 1978 e 2018 a população residente em Portugal aumentou cerca de 15%"

(0,75) b) Verifique, justificando com cálculos, se a afirmação seguinte é verdadeira ou falsa:

"Enquanto em 1978 existiam cerca de 2,5 vezes mais crianças com idades até aos 14 anos do que pessoas com mais de 65 anos, em 2018 a situação é radicalmente diferente, existindo pouco mais de 60 crianças nessa faixa etária por cada 100 indivíduos com 65 ou mais anos."

Exercício 2 (2,0 valores)

Na tabela abaixo pode ver a evolução das Dotações orçamentais públicas para investigação e desenvolvimento (I&D), a preços correntes.

Anos	Dotações orçamentais públicas para I&D (milhões de Euros)	PIB (2016=100)	Dotações orçamentais públicas para I&D (milhões de Euros) (Preços constantes)
2013	1.213,3	0,957	
2014	1.228,0	0,964	
2015	1.348,3	0,983	
2016	1.240,0	1,000	
2017	1.317,1	1,015	
2018	1.494,80	1,031	
2019	1.522,1	1,049	

Fontes de Dados: DGEEC/ME-MCTES; Fonte: PORDATA

(1,25) a) Deflacione a série de valores, utilizando o PIB como deflator, com base em 2016. Interprete o valor obtido em 2019.

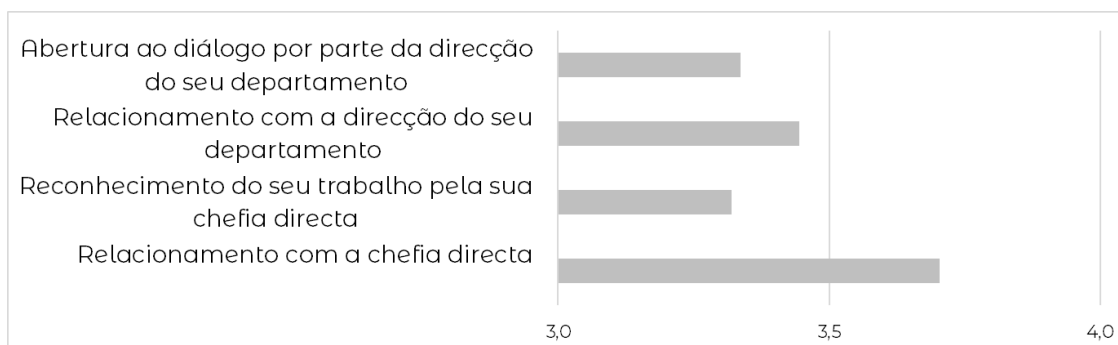
(0,75) b) Com os valores deflacionados, foi calculado o índice de base 2016, tendo sido apurado o valor de 104,6, em 2017. Interprete o seu significado.

Exercício 3 (7,5 valores)

A empresa ABC aplicou um questionário aos seus 128 colaboradores, com o objetivo de medir o nível de satisfação com diversos aspectos¹, nomeadamente com a chefia direta e com a direcção. A Figura 1 mostra um dos gráficos que surgiu no relatório apresentado à direcção.

¹ Escala de resposta utilizada: 1 -Nada satisfeito, 2-Pouco satisfeito, 3-Razoavelmente satisfeito, 4-Muito satisfeito e 5-Muitíssimo satisfeito.

Figura 1 – Nível médio de satisfação dos trabalhadores da empresa ABC com a chefia e com a direcção (1-nada satisfeito a 5-Muitissimo satisfeito)



(1,0) a) Na sua opinião, o gráfico encontra-se bem construído, de forma a possibilitar uma leitura comparativa dos resultados?

(1,5) b) Proponha alterações que considere serem importantes para melhorar esta visualização.

c) A tabela seguinte apresenta resultados das quatro variáveis anteriores.

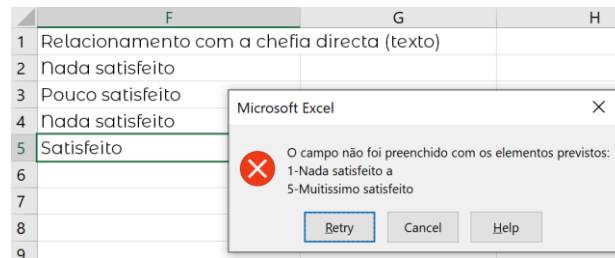
	Relacionamento com a chefia directa	Reconhecimento do seu trabalho pela sua chefia directa	Relacionamento com a direcção do seu departamento	Abertura ao diálogo por parte da direcção do seu departamento
Mean	3,70	3,32	3,44	3,34
Median	4	3	3	3
Standard Deviation	0,836	0,947	0,863	1,059
Count	74	85	126	128

(1,0) c1. Calcule e comente a taxa de não resposta à questão “Relacionamento com a chefia directa”.

(1,5) c2. Perante a presença do número de valores omissos nesta variável, qual seria a sua decisão, caso fosse o responsável por esta análise de dados? Escolha uma das seguintes opções e justifique a sua decisão.

- A. Ignorar os valores omissos e analisar somente as respostas válidas
- B. Efetuar uma imputação dos valores omissos
- C. Codificar os valores omissos de forma a analisar posteriormente os casos que não apresentaram resposta a esta variável.

- d) Quando as respostas foram introduzidas em Excel, pediu-se a um colaborador para preencher a tabela de dados com as respostas dadas à variável “Relacionamento com a chefia directa”, através da escrita das categorias. A dada altura o colaborador solicitou ajuda, porque surgiu um erro no ecrã e não estava a conseguir avançar:

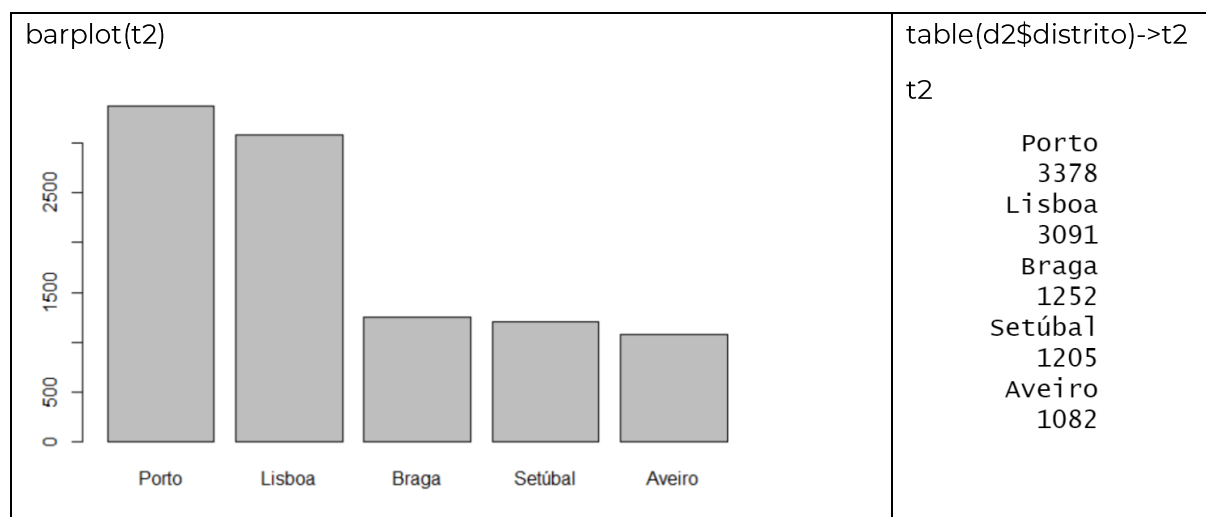


- (1,0) d1. Formule um texto com um breve esclarecimento para este colaborador.
- (1,5) d2. Descreva sucintamente os procedimentos a seguir para gerar este tipo de mensagem.

Exercício 4 (6,0 valores)

Os registos dos utentes com diabetes que acederam às unidades hospitalares, durante 2019, em Portugal, foram alvo de análise, com o objectivo de caracterização dos episódios de atendimento.

Uma das variáveis existente é o distrito em que ocorreu o atendimento. A análise foi realizada com a utilização do R. Apresentam-se em seguida os comandos utilizados para apurar o número de atendimentos deste tipo de utentes, por distrito (já filtrados para apenas 5), com a construção de uma tabela e respetivo gráfico de barras.



(1,5) a) Apresente o comando necessário para construir também uma tabela com os valores percentuais.

(1,5) b) Reescreva a função do gráfico, acrescentando:

- b1. Títulos aos dois eixos
- b2. Título do gráfico
- b3. Cor nas barras
- b4. Alteração do limite do eixo vertical
- b5. Atribuição do valor absoluto de atendimentos no distrito do Porto

(1,5) c) Como seria a linha de comando que permitisse construir novamente o gráfico sem “Aveiro”?

(1,5) d) Complete com os comandos de R necessários, em (A), (B) e (C), para escrever a linha de relatório, em R markdown:

É possível verificar que o maior número de atendimentos se verificou no distrito do

``r` (A) `, com `r` (B) ` utentes, correspondendo a `r` (C) ` %.`

Exercício 5 (3,0)

O **Anexo A** apresenta alguns resultados produzidos com o JAMOVl, relativos à percentagem de jovens até 25 anos (`inf_25_anos`), nas populações dos municípios de Portugal Continental. Os valores apresentam-se para o conjunto dos 278 municípios e distribuídos por regiões NUTSII²

Elabore uma descrição e interpretação dos resultados encontrados, identificando as figuras e/ou tabelas que mencione no seu texto.

² “Nomenclatura das Unidades Territoriais para Fins Estatísticos”, no nível II, Portugal Continental é formado por Norte, Centro, Área Metropolitana de Lisboa, Alentejo e Algarve.

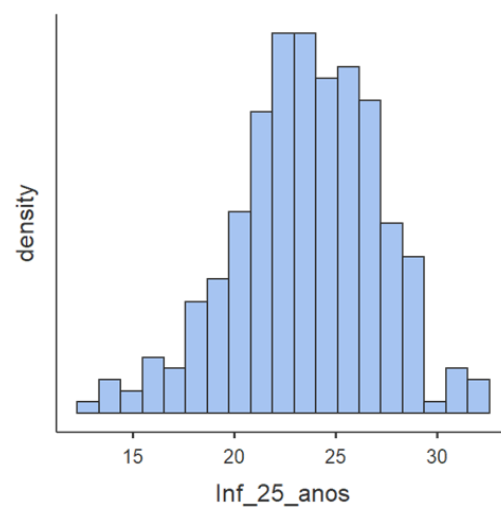
Anexo A

A1 -

	Inf_25_anos
n	278
Missing	0
Mean	23.6
Median	23.7
Mode	13.3 ^a
Standard deviation	3.53
Minimum	13.3
Maximum	32.5
Skewness	-0.309
Std. error skewness	0.146
Kurtosis	0.152
Std. error kurtosis	0.291
25th percentile	21.5
50th percentile	23.7
75th percentile	26.1

^a More than one mode exists, only the first is reported

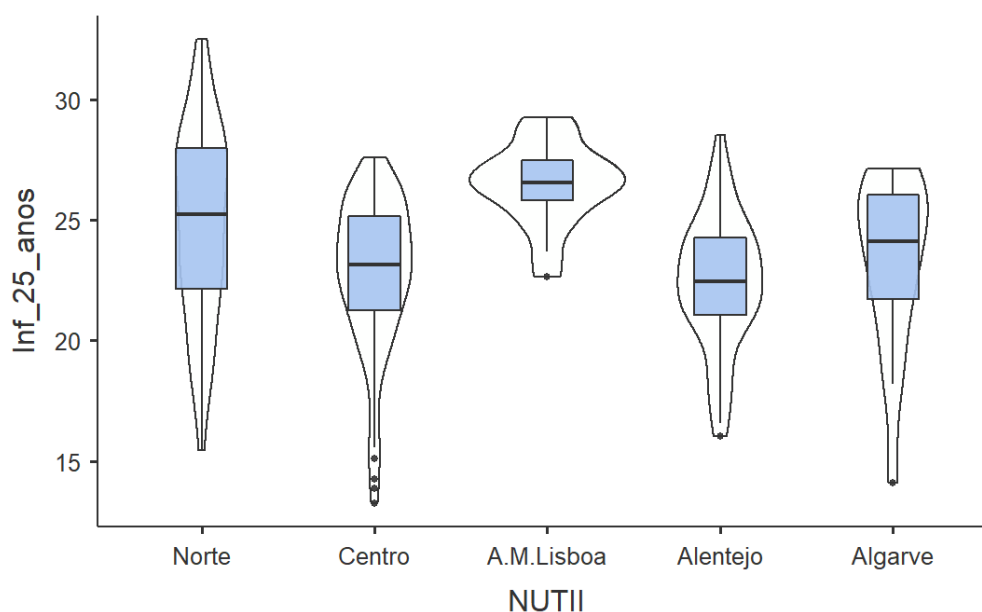
A2 -



A3 -

	NUTII	n	Mean	SD	SE
Inf_25_anos	Norte	86	24.8	3.99	0.430
	Centro	100	22.7	3.15	0.315
	A.M.Lisboa	18	26.5	1.70	0.402
	Alentejo	58	22.4	2.75	0.361
	Algarve	16	23.2	3.63	0.908

A4 -



A5 – Municípios com menor presença de jovens até aos 25 anos (%), por região NUTSII

NUTS II	Município	Valores extremos
Centro	Vila Velha de Ródão	13,3
	Pampilhosa da Serra	13,9
	Oleiros	14,3
	Penamacor	15,1
A.M.Lisboa	Lisboa	22,6
Alentejo	Gavião	16,0
Algarve	Alcoutim	14,1