



Estatística

Materiais de Apoio
Ensino Secundário



Maria Eugénia Graça Martins
Alexandra Rodrigues
Helder Martins
Nélia Amado
Susana Carreira





Título (E-book):
Estatística: Materiais de Apoio
Ensino Secundário

Autores:
Maria Eugénia Graça Martins, Alexandra Rodrigues, Helder Martins,
Nélia Amado, Susana Carreira

Coordenação:
Jaime Carvalho e Silva
Grupo de Trabalho do Desenvolvimento Curricular e Profissional em
Matemática para o Ensino Secundário (DCPMES)

Data:
Junho de 2024

**Direção-Geral da Educação, Ministério da Educação, Ciência e Inovação
Portugal**



ÍNDICE GERAL

NOTA INTRODUTÓRIA.....	1
PRINCÍPIOS DIDÁTICOS.....	3
“I HAVE A DREAM”	6
TAREFA	6
NOTAS PARA O PROFESSOR.....	13
PROPOSTA DE RESOLUÇÃO.....	19
A EVOLUÇÃO NA VIDA DA MULHER PORTUGUESA NOS ÚLTIMOS 50 ANOS.....	24
TAREFA	24
NOTAS PARA O PROFESSOR.....	33
PROPOSTA DE RESOLUÇÃO.....	36
DOÇURA OU TRAVESSURA (E A QUESTÃO DA DESIGUALDADE)	40
TAREFA	40
NOTAS PARA O PROFESSOR.....	43
PROPOSTA DE RESOLUÇÃO.....	46
CONSUMO DE ÁGUA	54
TAREFA	54
NOTAS PARA O PROFESSOR.....	58
PROPOSTA DE RESOLUÇÃO.....	61
QUANTO ESTICA UM ELÁSTICO?	68
TAREFA	68
NOTAS PARA O PROFESSOR.....	70
PROPOSTA DE RESOLUÇÃO.....	73
COMPREENDER UM ESTUDO SOBRE HIPERTENSÃO NOS JOVENS.....	79
TAREFA	79
NOTAS PARA O PROFESSOR.....	86
PROPOSTA DE RESOLUÇÃO.....	89
O COEFICIENTE DE CORRELAÇÃO LINEAR.....	93
PARA LER E COMPREENDER	93
TAREFA	98
NOTAS PARA O PROFESSOR.....	106
PROPOSTA DE RESOLUÇÃO.....	109
QUANTO MAIS POLUIÇÃO, MELHOR?.....	114
TAREFA	114
NOTAS PARA O PROFESSOR.....	117
PROPOSTA DE RESOLUÇÃO.....	119
NA HORA DE SUBSTITUIR OS PNEUS DO CARRO	121
TAREFA	121
NOTAS PARA O PROFESSOR.....	126
PROPOSTA DE RESOLUÇÃO.....	129
ÍNDICE BIG MAC	136
TRABALHO DE PROJETO.....	136
NOTAS PARA O PROFESSOR.....	137
DIAS QUENTES.....	141
TRABALHO DE PROJETO.....	141
NOTAS PARA O PROFESSOR.....	142
A CONCLUIR.....	147

Nota Introdutória

A Direção-Geral da Educação (DGE) tem vindo a conceber e a concretizar um conjunto de atividades destinadas a apoiar a generalização dos programas (Aprendizagens Essenciais) de Matemática para os 10º, 11º e 12º anos de escolaridade, designadamente nas disciplinas de Matemática A, Matemática B e Matemática dos Cursos Profissionais.

A presente *brochura* temática aborda os tópicos e subtópicos do tema de Estatística, que integra as Aprendizagens Essenciais de Matemática A, Matemática B e Matemática dos Cursos Profissionais, para o 10º ano de escolaridade.

Nesta *brochura* são apresentados:

- um conjunto de princípios didáticos, em linha com as Ações Estratégicas de Ensino do Professor mencionadas nas Aprendizagens Essenciais e em consonância com o Perfil dos Alunos à Saída da Escolaridade Obrigatória;
- uma coleção de tarefas de aprendizagem, concebidas para a aula de matemática, acompanhadas de notas específicas para o professor e de uma proposta de resolução detalhada e comentada;
- sugestões de Trabalho de Projeto, com indicações práticas;
- um índice remissivo, que identifica os tópicos e subtópicos abordados, os conceitos estatísticos, bem como os contextos subjacentes aos problemas e questões propostos nas tarefas.

A *brochura* foi elaborada no âmbito das atividades do Grupo de Trabalho (GT) do Desenvolvimento Curricular e Profissional em Matemática para o Ensino Secundário (DCPMES). Integram este GT os docentes e investigadores: Jaime Carvalho e Silva (Coordenador), Alexandra Rodrigues, Ana Breda, António Cardoso, António Domingos, Carlos Albuquerque, Cristina Cruchinho, Cristina Negra, Emanuel Martinho, Helder Manuel Martins, Hélia Jacinto, João Almiro, Luís Gabriel, Maria Eugénia Graça Martins, Maria Manuel Torres, Maria Teresa Santos, Nélia Amado, Nélida Filipe, Paulo Correia, Pedro Freitas, Pedro Macias Marques, Raúl Gonçalves, Rui Gonçalo Espadeiro e Susana Carreira.

As tarefas que compõem esta *brochura* aliam-se, em vários pontos, à visão consolidada no projeto ALEA – Ação Local Estatística Aplicada (www.alea.pt), sugerindo, designadamente, a consulta de recursos didáticos para o ensino da Estatística aí disponibilizados.

Algumas das tarefas aqui propostas foram experimentadas, discutidas e analisadas, com bons resultados, no contexto da formação inicial de professores de Matemática.

Os autores

Maria Eugénia Graça Martins

Alexandra Rodrigues

Helder Martins

Nélia Amado

Susana Carreira

O coordenador do GT DCPMES

Jaime Carvalho e Silva

Princípios Didáticos

Pensamento estatístico e literacia estatística

A Estatística utiliza formas de pensamento específicas e o desenvolvimento do pensamento estatístico, no ensino secundário, deve incluir abordagens exploratórias de dados reais, em que se detetam padrões e tendências, se identificam modelos e se reconhecem casos excecionais num conjunto de dados. É importante formar cidadãos estatisticamente instruídos que sejam capazes de compreender ideias e conceitos estatísticos num mundo cada vez mais sustentado em dados. As atividades realistas, que recorrem à produção e/ou recolha de dados autênticos visam uma aproximação à prática estatística profissional, dando ênfase ao raciocínio e à argumentação, incluindo: i) compreensão do contexto dos dados; ii) exploração dos dados; iii) utilização de resultados estatísticos como evidência para apoiar argumentos de decisão e/ou soluções informadas.

Nesta *brochura* são privilegiados contextos reais e situações relevantes do mundo atual, bem como o recurso a dados autênticos – disponíveis em bases de dados nacionais e internacionais. É igualmente estimulada a produção de argumentos e reflexões, no contexto real dos problemas abordados, designadamente, a propósito de relatórios e notícias veiculados pelos meios de comunicação social ou de estudos empíricos. Também nesse sentido, podem encontrar-se várias oportunidades de trabalho e de diálogo interdisciplinar nos materiais aqui propostos.

A tecnologia como parte integrante da investigação estatística

A tecnologia computacional, atualmente disponível, permite formas expeditas de organizar dados, manipular grandes quantidades de dados, tratar diferentes tipos de dados, visualizar e representar dados de múltiplas formas. O recurso à tecnologia torna-se parte integrante da análise de dados e constitui uma das características essenciais da tendência protagonizada pela ciência de dados: uma combinação multidisciplinar de computação, estatística e modelação matemática. Tirar partido de software específico, como a folha de cálculo ou o GeoGebra, ou utilizar uma calculadora gráfica, elimina barreiras, nomeadamente de cálculo rotineiro. O foco é colocado na experimentação, no raciocínio, na representação, na visualização e na

interpretação. Isto significa que os alunos poderão envolver-se em atividades autênticas de análise e interpretação de dados reais. O desenvolvimento do pensamento estatístico está intimamente relacionado com a aquisição de uma competência tecno-matemática, que torna os alunos capazes de identificar e selecionar as ferramentas e oportunidades adequadas aos objetivos das suas investigações estatísticas.

Todas as tarefas e questões propostas nesta *brochura* requerem o recurso sistemático à tecnologia, com destaque para a folha de cálculo enquanto ferramenta que oferece as funcionalidades requeridas e incentiva o raciocínio necessário para a exploração e representação de dados autênticos (nomeadamente importados de ficheiros disponíveis online).

Visualização de conceitos estatísticos

A visualização de conceitos estatísticos está associada, não apenas à possibilidade de criação de representações dinâmicas e interativas, com recurso à tecnologia, mas também ao facto de os alunos serem desafiados a realizar investigação estatística, pondo em prática conceitos e definições num contexto concreto.

A construção de múltiplas representações gráficas para descrever e interpretar distribuições de dados, univariados e multivariados, permite obter modelos visuais de situações reais, favorecendo o pensamento estatístico e a compreensão conceptual, isto é, facilitando a atribuição de sentido aos conceitos estatísticos. Em articulação com as aprendizagens desenvolvidas no ensino básico, os alunos do ensino secundário continuarão a ser incentivados a construir e interpretar gráficos e diagramas de diversos tipos e a avaliar a adequação de tais representações aos propósitos do seu estudo estatístico.

As representações gráficas permitem a visualização de conceitos estatísticos essenciais, como a variabilidade, a incerteza, a concentração, ou a associação entre variáveis. Na presente *brochura*, apela-se frequentemente à representação gráfica e à comunicação matemática de resultados e respetiva interpretação.

Contextos de aprendizagem

Os alunos deverão ter oportunidades de reconhecer e valorizar o papel da Estatística na compreensão de situações do mundo real – sejam estas de carácter social, económico, ambiental ou outras – incluindo a consciência de que problemas complexos da atualidade são abordados a partir da obtenção e análise de grandes quantidades de dados. A interpretação e a produção de informação estatística são competências essenciais para uma cidadania ativa e informada. Os contextos devem ser parte integrante da análise e interpretação de dados reais, de tal modo que os alunos, através do conhecimento estatístico, consigam estabelecer relações significativas com o mundo em seu redor e com a realidade social e cultural em que participam. Isto implica, como se advoga nesta *brochura*, que o aluno seja convidado

a tomar uma posição ou uma decisão, a contextualizar os seus resultados e a comunicá-los, usando argumentos baseados em evidência estatística. Trata-se, portanto, de promover e desenvolver o sentido crítico dos alunos, enquanto cidadãos produtores e consumidores de dados.

“I HAVE A DREAM”

TAREFA



O documento em anexo contém o discurso, traduzido para língua portuguesa¹, proferido pelo norte americano Martin Luther King, em 28 de Agosto de 1963, no Lincoln Memorial, em Washington D.C.

Em 1964, com 35 anos de idade, este defensor dos direitos humanos foi galardoado com o Prémio Nobel da Paz. (Para mais informações, consultar: <https://www.nobelprize.org/prizes/peace/1964/king/biographical/>)

Um discurso político tem características linguísticas que o tornam distinto de outros géneros de discurso, por exemplo, o uso de metáforas, metonímias, anáforas, etc. Um dos grandes objetivos do discurso político é o de persuadir o público dos ideais ou visões do orador.

A Estatística tem um papel importante na análise linguística dos textos, permitindo, em particular, estabelecer padrões de comparação entre estilos de diferentes autores ou géneros de discursos, por meio da análise dos elementos (palavras, pontuação, etc.) que compõem o texto.

Nesta tarefa, iremos analisar este discurso, que ficou conhecido na história mundial, estabelecendo como objetivo a seguinte questão:

Qual o número médio de letras, ou tamanho médio, das palavras do célebre discurso (em Português) de Martin Luther King, “I have a dream”?

¹ A tradução foi obtida de: <https://www.argnet.pt/portal/discursos/agosto05.html>

1. Faz uma leitura do discurso e arrisca uma estimativa (sem fazeres contagens) para o número médio de letras das palavras do discurso.
2. Considera as palavras que formam a primeira frase do discurso e calcula a média do número de letras dessas palavras. Será esse valor uma boa estimativa para o tamanho médio das palavras do discurso? Justifica.
3. Como proceder para chegar a uma estimativa do tamanho médio, sem contar as letras de todas as palavras do discurso? Explica a tua resposta.
4. Para efetuares uma análise estatística do discurso, começa por identificar a população, a unidade estatística e a(s) variável(variáveis) em estudo.

4.1. Tens à tua disposição um ficheiro de Excel com o nome "I HAVE A DREAM". Nesse ficheiro, há uma folha intitulada **Amostras** e outra intitulada **Análise**. Abre o ficheiro na folha **Amostras**, onde podes encontrar, nas três primeiras colunas: a *Etiqueta* que dá a posição de cada palavra no discurso, a *População* que contém cada uma das palavras do discurso, o *Número de Letras* que mostra o tamanho de cada palavra.

Na célula assinalada a amarelo (ver figura), introduz o valor 10 para dimensão da amostra. Obténs automaticamente 3 amostras aleatórias simples de dimensão 10, selecionadas da população.

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
1	Etiqueta	População	Nº de letras	Dimensão da amostra	Amostra 1												
2					Etiqueta	Palavras	Tam.		Etiqueta	Palavras	Tam.		Etiqueta	Palavras	Tam.		
3	1	Há	2		#CÁLC!				#CÁLC!				#CÁLC!				
4	2	cem	3														
5	3	anos	4														
6	4	um	2														
7	5	grande	6														
8	6	americano	9														
9	7	sob	3														
10	8	cuja	4														
11	9	sombra	6														
12	10	simbólica	9														
13	11	nos	3														
14	12	encontramos	11														
15	13	assinava	8														
16	14	o	1														
17	15	Proclamação	11														
18	16	da	2														

Em seguida, copia os dados de cada amostra e cola os valores na folha **Análise** (Usa **colar especial** e escolhe a opção "**colar valores**").

Nota: Sempre que é executada uma ação no Excel, são recalculados automaticamente os valores gerados aleatoriamente, o que faz alterar a composição da amostra já extraída. Por isso, quando se usa "colar valores" (isto é, sem fórmulas), os valores colados permanecem inalterados.

Na folha **Análise**, obtém as seguintes medidas, para cada amostra: média, mediana, mínimo, máximo e amplitude. Compara os resultados que obtiveste nas três amostras. Consegues dizer qual das três médias será a que melhor representa o valor médio da população em estudo?

Nota: As seguintes fórmulas fazem o cálculo das medidas pretendidas.

=MÉDIA(Célula Inicial:Célula Final)

=MED(Célula Inicial:Célula Final)

=MÍNIMO(Célula Inicial:Célula Final)

=MÁXIMO(Célula Inicial:Célula Final)

4.2. Para obteres uma representação gráfica da variabilidade que existe nas várias amostras, constrói no Excel um *diagrama em caixa de bigodes* para cada amostra. Compara os diagramas e tira conclusões quanto à variabilidade.

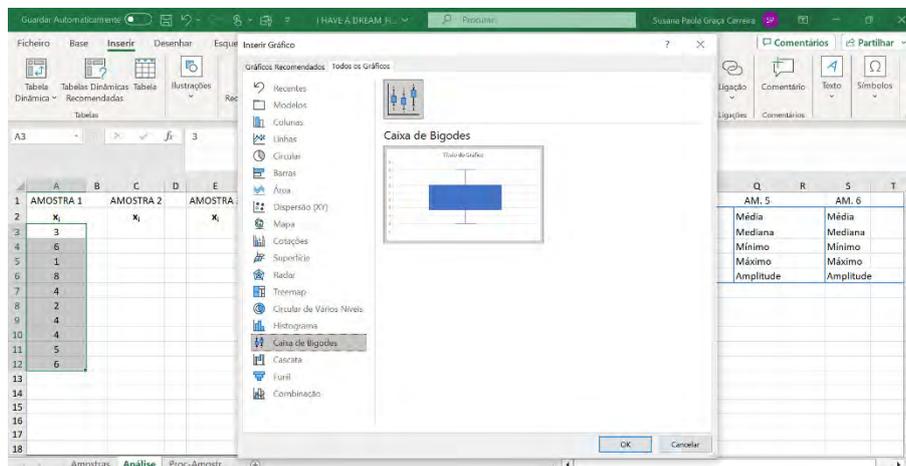
Nota: Para criares um diagrama em caixa de bigodes, seleciona o conjunto de dados que queres representar e vai ao menu INSERIR.

Escolhe: Inserir Gráficos.

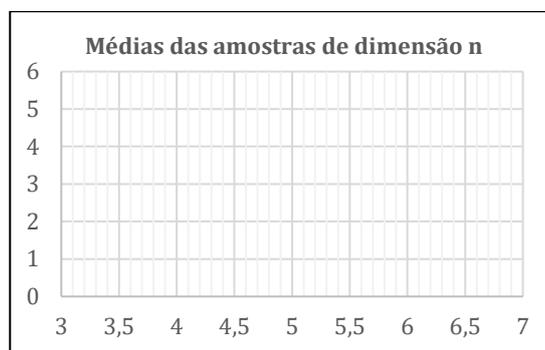
Seleciona a aba: Todos os Gráficos.

Encontra a opção Caixa de Bigodes. Escolhendo essa opção, tens imediatamente uma pré-visualização do gráfico. (Ver figura).

Clica em OK e o diagrama aparece na folha do Excel.



4.3. Fazendo a recolha das médias obtidas por todos os alunos nas suas amostras, todos irão construir, coletivamente, um *diagrama de pontos*. (Para a construção do diagrama de pontos, poderá ser usado um formato como o sugerido abaixo).



Cada aluno, ou cada grupo, marcará no gráfico as médias das suas amostras, arredondadas às décimas. O eixo horizontal corresponde ao valor da média da amostra e o vertical corresponde ao número de vezes que ocorreu essa média.

Por exemplo, se há 4 alunos da turma que obtiveram média 5 nas suas amostras, estas são representadas pelos pontos de coordenadas (5,1), (5,2), (5,3), (5,4).

Observa o diagrama de pontos obtido e retira conclusões quanto à variabilidade dos resultados.

5. Como será o comportamento da distribuição das médias se as amostras forem de maior dimensão? Apresentarão a mesma variabilidade entre si? Justifica.

6. Altera o valor da dimensão da amostra para 50 e repete a **análise feita nos itens 4.1 a 4.3**.

7. A partir do diagrama de pontos coletivo, após discussão com os teus colegas, sugere um valor consensual como estimativa para o tamanho médio das palavras do discurso. Calcula depois o tamanho médio das palavras, considerando todas as unidades observacionais que constituem a população, e compara-o com o valor que foi sugerido pela turma.

8. Escreve uma pequena composição (cerca de 15-20 linhas) em que explicas de forma breve as análises estatísticas efetuadas e as conclusões a que chegaste quanto à influência da dimensão da amostra na obtenção de informação útil acerca do *tamanho médio* das palavras do discurso “I have a dream”, de Martin Luther King, ou de uma forma mais geral, quando se pretende estimar o *valor médio* de uma população, utilizando a *média* de uma amostra.

ANEXO

Texto do discurso "I have a dream", de Martin Luther King

Há cem anos, um grande americano, sob cuja sombra simbólica nos encontramos, assinava a Proclamação da Emancipação. Esse decreto fundamental foi como um raio de luz de esperança para milhões de escravos negros que tinham sido marcados a ferro nas chamas de uma vergonhosa injustiça. Veio como uma aurora feliz para terminar a longa noite do cativo. Mas, cem anos mais tarde, devemos enfrentar a realidade trágica de que o Negro ainda não é livre.

Cem anos mais tarde, a vida do Negro é ainda lamentavelmente dilacerada pelas algemas da segregação e pelas correntes da discriminação. Cem anos mais tarde, o Negro continua a viver numa ilha isolada de pobreza, no meio de um vasto oceano de prosperidade material. Cem anos mais tarde, o Negro ainda definha nas margens da sociedade americana, estando exilado na sua própria terra.

Por isso, encontramos-nos aqui hoje para dramaticamente mostrarmos esta extraordinária condição. Num certo sentido, viemos à capital do nosso país para descontar um cheque. Quando os arquitetos da nossa república escreveram as magníficas palavras da Constituição e da Declaração de Independência, estavam a assinar uma promissória de que cada cidadão americano se tornaria herdeiro.

Este documento era uma promessa de que todos os homens veriam garantidos os direitos inalienáveis à vida, à liberdade e à procura da felicidade. É óbvio que a América ainda hoje não pagou tal promissória no que concerne aos seus cidadãos de cor. Em vez de honrar este compromisso sagrado, a América deu ao Negro um cheque sem cobertura; um cheque que foi devolvido com a seguinte inscrição: "saldo insuficiente". Porém nós recusamo-nos a aceitar a ideia de que o banco da justiça esteja falido. Recusamo-nos a acreditar que não exista dinheiro suficiente nos grandes cofres de oportunidades deste país.

Por isso viemos aqui cobrar este cheque - um cheque que nos dará quando o recebermos as riquezas da liberdade e a segurança da justiça. Também viemos a este lugar sagrado para lembrar à América da clara urgência do agora. Não é o momento de se dedicar à luxúria do adiamento, nem para se tomar a pílula tranquilizante do gradualismo. Agora é tempo de tornar reais as promessas da Democracia. Agora é o tempo de sairmos do vale escuro e desolado da segregação para o iluminado caminho da justiça racial. Agora é tempo de abrir as portas da oportunidade para todos os filhos de Deus. Agora é tempo para retirar o nosso país das areias movediças da injustiça racial para a rocha sólida da fraternidade.

Seria fatal para a nação não levar a sério a urgência do momento e subestimar a determinação do Negro. Este sufocante verão do legítimo descontentamento do Negro não passará até que chegue o revigorante Outono da liberdade e igualdade. 1963 não é um fim, mas um começo. Aqueles que creem que o Negro precisava só de desabafar, e que a partir de agora ficará sossegado, irão acordar sobressaltados se o País regressar à sua vida de sempre. Não haverá tranquilidade nem descanso na América até que o Negro tenha garantido todos os seus direitos de cidadania.

Os turbilhões da revolta continuarão a sacudir as fundações do nosso País até que desponte o luminoso dia da justiça. Existe algo, porém, que devo dizer ao meu povo que se encontra no caloroso limiar que conduz ao palácio da justiça. No percurso de ganharmos o nosso legítimo lugar não devemos ser culpados de atos errados. Não tentemos satisfazer a sede de liberdade bebendo da taça da amargura e do ódio.

Temos de conduzir a nossa luta sempre no nível elevado da dignidade e disciplina. Não devemos deixar que o nosso protesto realizado de uma forma criativa degenere na violência física. Teremos de nos erguer uma e outra vez às alturas majestosas para enfrentar a força física com a força da consciência.

Esta maravilhosa nova militância que engolfou a comunidade negra não nos deve levar a desconfiar de todas as pessoas brancas, pois muitos dos nossos irmãos brancos, como é claro pela sua presença

aqui, hoje, estão conscientes de que os seus destinos estão ligados ao nosso destino, e que sua liberdade está intrinsecamente ligada à nossa liberdade.

Não podemos caminhar sozinhos. À medida que caminhamos, devemos assumir o compromisso de marcharmos em frente. Não podemos retroceder. Há quem pergunte aos defensores dos direitos civis: "Quando é que ficarão satisfeitos?" Não estaremos satisfeitos enquanto o Negro for vítima dos incontáveis horrores da brutalidade policial. Não poderemos estar satisfeitos enquanto os nossos corpos, cansados das fadigas da viagem, não conseguirem ter acesso a um lugar de descanso nos motéis das estradas e nos hotéis das cidades. Não poderemos estar satisfeitos enquanto a mobilidade fundamental do Negro for passar de um gueto pequeno para um maior. Nunca poderemos estar satisfeitos enquanto um Negro no Mississípi não pode votar e um Negro em Nova Iorque achar que não há nada pelo qual valha a pena votar. Não, não, não estamos satisfeitos, e só ficaremos satisfeitos quando a justiça correr como a água e a retidão como uma poderosa corrente.

Sei muito bem que alguns de vocês chegaram aqui após muitas dificuldades e tribulações. Alguns de vocês saíram recentemente de pequenas celas de prisão. Alguns de vocês vieram de áreas onde a vossa procura da liberdade vos deixou marcas provocadas pelas tempestades da perseguição e sofrimentos provocados pelos ventos da brutalidade policial. Vocês são veteranos do sofrimento criativo. Continuem a trabalhar com a fé de que um sofrimento injusto é redentor.

Voltem para o Mississípi, voltem para o Alabama, voltem para a Carolina do Sul, voltem para a Geórgia, voltem para a Luisiana, voltem para os bairros de lata e para os guetos das nossas modernas cidades, sabendo que, de alguma forma, esta situação pode e será alterada. Não nos embrenhemos no vale do desespero.

Digo-lhes, hoje, meus amigos, que apesar das dificuldades e frustrações do momento, ainda tenho um sonho. É um sonho profundamente enraizado no sonho americano.

Tenho um sonho que um dia esta nação levantar-se-á e viverá o verdadeiro significado da sua crença: "Consideramos estas verdades como evidentes por si mesmas, que todos os homens são criados iguais".

Tenho um sonho que um dia nas montanhas rubras da Geórgia os filhos de antigos escravos e os filhos de antigos proprietários de escravos poderão sentar-se à mesa da fraternidade.

Tenho um sonho que um dia o estado do Mississípi, um estado deserto, sufocado pelo calor da injustiça e da opressão, será transformado num oásis de liberdade e justiça.

Tenho um sonho que meus quatro pequenos filhos viverão um dia numa nação onde não serão julgados pela cor da sua pele, mas pela qualidade do seu carácter.

Tenho um sonho, hoje.

Tenho um sonho que um dia o estado de Alabama, cujos lábios do governador atualmente pronunciam palavras de... e recusa, seja transformado numa condição onde pequenos rapazes negros, e raparigas negras, possam dar-se as mãos com outros pequenos rapazes brancos, e raparigas brancas, caminhando juntos, lado a lado, como irmãos e irmãs.

Tenho um sonho, hoje.

Tenho um sonho que um dia todos os vales serão elevados, todas as montanhas e encostas serão niveladas, os lugares ásperos serão polidos, e os lugares tortuosos serão endireitados, e a glória do Senhor será revelada, e todos os seres a verão, conjuntamente.

Esta é nossa esperança. Esta é a fé com a qual regresso ao Sul. Com esta fé seremos capazes de retirar da montanha do desespero uma pedra de esperança. Com esta fé poderemos transformar as dissonantes discórdias de nossa nação numa bonita e harmoniosa sinfonia de fraternidade. Com esta fé poderemos trabalhar juntos, rezar juntos, lutar juntos, ir para a prisão juntos, ficarmos juntos em posição de sentido pela liberdade, sabendo que um dia seremos livres.

Esse será o dia quando todos os filhos de Deus poderão cantar com um novo significado: "O meu país é teu, doce terra de liberdade, de ti eu canto. Terra onde morreram os meus pais, terra do orgulho dos peregrinos, que de cada localidade ressoe a liberdade".

E se a América quiser ser uma grande nação isto tem que se tornar realidade. Que a liberdade ressoe então dos prodigiosos cabeços do Novo Hampshire. Que a liberdade ressoe das poderosas montanhas de Nova Iorque. Que a liberdade ressoe dos elevados Alleghenies da Pensilvânia!

Que a liberdade ressoe dos cumes cobertos de neve das montanhas Rochosas do Colorado!

Que a liberdade ressoe dos picos curvos da Califórnia!

Mas não só isso; que a liberdade ressoe da Montanha de Pedra da Geórgia!

Que a liberdade ressoe da Montanha Lookout do Tennessee!

Que a liberdade ressoe de cada Montanha e de cada pequena elevação do Mississípi.

Que de cada localidade, a liberdade ressoe.

Quando permitirmos que a liberdade ressoe, quando a deixarmos ressoar de cada vila e cada aldeia, de cada estado e de cada cidade, seremos capazes de apressar o dia em que todos os filhos de Deus, negros e brancos, judeus e gentios, protestantes e católicos, poderão dar-se as mãos e cantar as palavras da antiga canção negra: "Liberdade finalmente! Liberdade finalmente! Louvado seja Deus, Todo-Poderoso, estamos livres, finalmente!"

NOTAS PARA O PROFESSOR

A exploração da tarefa “**I have a dream**” é adequada às Aprendizagens Essenciais de Matemática A, Matemática B, MACS e Matemática dos Cursos Profissionais.

A tarefa tem um grande foco no desenvolvimento da Comunicação Matemática e constitui um exemplo de uma aplicação autêntica da Estatística à análise linguística de textos, promovendo o recurso sistemático à tecnologia (em particular o uso da folha de cálculo) e o desenvolvimento do espírito crítico.

Tema: Estatística

Tópicos e subtópicos:

Problema estatístico

- Variabilidade

População, amostra e variável

Dados univariados

- Dados quantitativos discretos

- Organização de dados

- Medidas de localização

- Medidas de dispersão

Objetivos de aprendizagem:

- Reconhecer a variabilidade como um conceito chave de um problema estatístico.
- Identificar num estudo estatístico, população, amostra e a(s) característica(s) a estudar, que se designa(m) por variável(variáveis).
- Reconhecer as fases de um procedimento estatístico: Produção ou aquisição de dados; Organização e representação de dados; Interpretação tendo por base as representações obtidas.
- Reconhecer os métodos existentes para a seleção de amostras, no sentido de que estas sejam representativas das populações subjacentes, e de modo a evitar amostras enviesadas cujo estudo levaria a inferir conclusões erradas para as populações.
- Intuir que os problemas estatísticos, em que se recorre a amostras para inferir para a população subjacente, não têm uma solução matemática única que se possa exprimir como verdadeiro ou falso.
- Selecionar representações gráficas adequadas para cada tipo de dados identificando vantagens/inconvenientes, lembrando os diagramas de extremos e quartis.
- Interpretar as medidas de localização: média (\bar{x}), mediana (Me), moda(s) (Mo) e percentis (quartis como caso especial) na caracterização da distribuição dos dados, relacionando-as com as representações gráficas obtidas.
- Interpretar as medidas de dispersão, amplitude, amplitude interquartil na caracterização da distribuição dos dados, relacionando-as com as representações gráficas obtidas.

Duração prevista: Duas aulas de 50 minutos.

Materiais e recursos:

- Computador, telemóvel.
- Internet, folha de cálculo.

Estratégias de implementação:

O discurso de Martin Luther King, “I have a dream”, dá o mote a uma tarefa estatística que permite selecionar múltiplas amostras de diferentes dimensões, usando a folha de cálculo, que pode ser disponibilizada pronta a funcionar de forma automática ou ser construída pelos alunos. Em anexo, incluem-se algumas indicações de apoio à construção da folha de cálculo, nomeadamente sobre as principais funções usadas para obter amostras aleatórias simples.

Recomendamos que numa fase anterior à aula, o professor disponibilize o texto do discurso, para que os alunos possam proceder à sua leitura atenta.

No início da aula, os alunos poderão ser convidados a falar sobre o que desejam destacar da leitura do discurso e a comentar o texto. Em paralelo, poderá promover-se a interdisciplinaridade com o professor de Português, de História ou de Área de Integração. Entre outros aspetos, poderá ser abordado o problema sistémico de racismo que se coloca em muitos países do mundo.

Prevê-se que a tarefa seja resolvida em pares, podendo considerar-se outras opções caso não exista um computador disponível por cada par. A tarefa permite a revisão de alguns conceitos estatísticos estudados no ensino básico. A resolução da tarefa deverá dar oportunidade a que os alunos se apropriem dos conceitos estatísticos subjacentes.

Na questão 3, é importante promover a discussão sobre o cuidado a ter com a recolha da amostra, exemplificando com a recolha de amostras por conveniência ou resposta voluntária, que dão origem a amostras enviesadas. Deve ser claro para os alunos que os resultados do estudo da amostra só poderão ser generalizados para a população, se a amostra for representativa dessa população. Poderá ser referida a existência de processos apropriados para a seleção das amostras de forma a garantir a aleatoriedade e a representatividade da população subjacente, nomeadamente os processos de amostragem aleatória. É possível recorrer ao Excel para selecionar várias amostras aleatórias de dimensões diferentes.

Na questão 4.2, poderá surgir a noção de outlier na representação gráfica em caixa de bigodes, que é nova para os alunos, pelo que o professor deverá explicar o conceito, referindo porque é que um mesmo valor não é, necessariamente, um outlier em todas as amostras consideradas.

A pergunta 4.3. é promotora de uma discussão coletiva, a partir da construção de um diagrama de pontos (por ex., no quadro, no GeoGebra ou no Excel) onde serão representados os valores das médias das amostras, da mesma dimensão. A partir desta construção deverá dinamizar-se uma análise coletiva sobre a variabilidade das amostras consideradas e tirar algumas conclusões.

Na última fase da tarefa, as questões 7 e 8 deverão ser corrigidas coletivamente e sugere-se que seja promovida uma discussão crítica dos resultados em grande grupo. Na discussão global, os alunos compararão as diferentes amostras que usaram e refletirão sobre a importância da dimensão da amostra. No final da aula, os alunos refletem individualmente sobre o seu trabalho e o que aprenderam.

Esta tarefa pode ser adaptada pelo docente às características da escola, do meio sociocultural envolvente e do grupo turma.

Dificuldades esperadas e ações do professor:

Os alunos podem revelar algumas dificuldades, nomeadamente:

- No uso da Folha de Cálculo. O professor deve estar disponível para ajudar os alunos individualmente, sendo recomendável organizar os alunos em pares para promover a interajuda.
- Em recordar e retomar algumas noções estatísticas estudadas no ensino básico, que permitam interpretar as medidas de localização, de dispersão e a representação gráfica em caixa de bigodes. Esta tarefa será um excelente ponto de partida para rever estes conceitos.

O professor, sempre que necessário, deve rever os conceitos estatísticos de anos anteriores e ajudar os alunos na utilização da tecnologia (folha de cálculo e/ou calculadora gráfica), dando sugestões ou colocando questões que contribuam para o desenvolvimento do trabalho.

Avaliação:

A avaliação poderá ter um carácter formativo.

Sugere-se que sejam recolhidas as respostas individuais dos alunos à questão 8 e que o docente dê feedback escrito sobre as mesmas. Esta prática poderá ser importante para a organização da aula seguinte, pois permitirá analisar o grau de compreensão dos alunos em relação à influência da dimensão da amostra na análise de dados.

Possíveis Aprofundamentos:

Usando as mesmas funções que estiveram na base da construção feita na folha de cálculo, poderão ser obtidas diversas amostras desde que se disponha da listagem dos elementos da população. Exemplos de populações das quais pode ser útil a seleção de amostras são: os sócios de um clube desportivo, os membros de uma ordem profissional, os elementos um agrupamento escolar, seja no que diz respeito aos alunos, ao pessoal docente ou ao pessoal discente, etc.

A metodologia aqui seguida para obter o tamanho médio das palavras de um dado texto, pode ser utilizada para estudar as distribuições do tamanho das palavras de textos de autores diferentes, com o objetivo de comparar os seus estilos literários.

ANEXO

Indicações para obter amostras aleatórias simples no Excel

a) Começar por fornecer aos alunos uma tabela em Excel onde estão listadas todas as palavras numeradas de 1 a 1513 (nas colunas A e B), de que se apresenta a seguir um pequeno excerto:

	A	B	C	D
1	Etiqueta	População		
2				
3	1	Há		
4	2	cem		
5	3	anos		
6	4	um		
7	5	grande		
8	6	americano		
9	7	sob		
10	8	cuja		
11	9	sombra		
12	10	simbólica		
13	11	nos		
14	12	encontramos		

	A	B	C	D
1502	1500	canção		
1503	1501	negra		
1504	1502	Liberdade		
1505	1503	finalmente		
1506	1504	Liberdade		
1507	1505	finalmente		
1508	1506	Louvado		
1509	1507	seja		
1510	1508	Deus		
1511	1509	Todo		
1512	1510	Poderosa		
1513	1511	estamos		
1514	1512	livres		
1515	1513	finalmente		
1516				
1517				

b) Na coluna C, utilizar a função **ALEATÓRIOENTRE(1;1513)**(ver Notas 1 e 4) para gerar tantos números aleatórios entre 1 e 1513, quantos a dimensão da amostra a considerar, replicando a fórmula até à linha necessária (de acordo com a dimensão da amostra). Por exemplo, o resultado de gerar 10 números aleatórios, conduziu-nos às seguintes etiquetas:

	A	B	C
1	Etiqueta	População	Nºs Gerados
2			
3	1	Há	=ALEATÓRIOENTRE(1;1513)
4	2	cem	=ALEATÓRIOENTRE(1;1513)
5	3	anos	=ALEATÓRIOENTRE(1;1513)
6	4	um	=ALEATÓRIOENTRE(1;1513)
7	5	grande	=ALEATÓRIOENTRE(1;1513)
8	6	americano	=ALEATÓRIOENTRE(1;1513)
9	7	sob	=ALEATÓRIOENTRE(1;1513)
10	8	cuja	=ALEATÓRIOENTRE(1;1513)
11	9	sombra	=ALEATÓRIOENTRE(1;1513)
12	10	simbólica	=ALEATÓRIOENTRE(1;1513)
13	11	nos	
14	12	encontramos	
15	13	assinava	
16	14	a	
17	15	Proclamação	

	A	B	C
1	Etiqueta	População	Nºs Gerados
2			
3	1	Há	895
4	2	cem	887
5	3	anos	677
6	4	um	571
7	5	grande	726
8	6	americano	1280
9	7	sob	575
10	8	cuja	620
11	9	sombra	125
12	10	simbólica	434
13	11	nos	
14	12	encontramos	
15	13	assinava	
16	14	a	
17	15	Proclamação	

c) Considerar para unidades estatísticas da amostra as palavras cujas etiquetas tenham sido seleccionadas. Para encontrar, na população, essas palavras, aconselha-se a utilizar na coluna D a função **PROCV** (ver Nota 2) que procura um valor de uma tabela. No nosso caso, a tabela onde se pretende fazer a pesquisa corresponde às células A3:B1515, que contêm a população. A fórmula a introduzir é a seguinte: **PROCV(C3;\$A\$3:\$B\$1515;2)**.

	A	B	C	D
1	Etiqueta	População	Nºs Gerados	
2				
3	1	Há	895	=PROCV(C3;\$A\$3:\$B\$1515;2)
4	2	cem	887	=PROCV(C4;\$A\$3:\$B\$1515;2)
5	3	anos	677	=PROCV(C5;\$A\$3:\$B\$1515;2)
6	4	um	571	=PROCV(C6;\$A\$3:\$B\$1515;2)
7	5	grande	726	=PROCV(C7;\$A\$3:\$B\$1515;2)
8	6	americano	1280	=PROCV(C8;\$A\$3:\$B\$1515;2)
9	7	sob	575	=PROCV(C9;\$A\$3:\$B\$1515;2)
10	8	cuja	620	=PROCV(C10;\$A\$3:\$B\$1515;2)
11	9	sombra	125	=PROCV(C11;\$A\$3:\$B\$1515;2)
12	10	simbólica	434	=PROCV(C12;\$A\$3:\$B\$1515;2)
13	11	nos		
14	12	encontramos		
15	13	assinava		
16	14	a		
17	15	Proclamação		

As palavras que constituem a amostra (cujas etiquetas correspondem aos números gerados aleatoriamente) são as que surgem na coluna D:

	A	B	C	D	E
1	Etiqueta	População	Nºs Gerados		
2					
3	1	Há	895	sofrimento	
4	2	cem	887	ventos	
5	3	anos	677	ligados	
6	4	um	571	satisfazer	
7	5	grande	726	enquanto	
8	6	americano	1280	sabendo	
9	7	sob	575	liberdade	
10	8	cuja	620	outra	
11	9	sombra	125	Negro	
12	10	simbólica	434	do	
13	11	nos			
14	12	encontramos			
15	13	assinava			
16	14	a			
17	15	Proclamação			
18	16	da			

d) Finalmente, para obter os dados, contar o número de letras das palavras da amostra. Esta contagem é facilitada pela utilização da função **NÚM.CARACT** (ver Nota 3) que será inserida na coluna E.

	A	B	C	D	E
1	Etiqueta	População	Nºs Gerados		
2					
3	1	Há	895	sofrimento	=NÚM.CARAT(D3)
4	2	cem	887	ventos	=NÚM.CARAT(D4)
5	3	anos	677	ligados	=NÚM.CARAT(D5)
6	4	um	571	satisfazer	=NÚM.CARAT(D6)
7	5	grande	726	enquanto	=NÚM.CARAT(D7)
8	6	americano	1280	sabendo	=NÚM.CARAT(D8)
9	7	sob	575	liberdade	=NÚM.CARAT(D9)
10	8	cuja	620	outra	=NÚM.CARAT(D10)
11	9	sombra	125	Negro	=NÚM.CARAT(D11)
12	10	simbólica	434	do	=NÚM.CARAT(D12)
13	11	nos			
14	12	encontramos			
15	13	assinava			
16	14	a			
17	15	Proclamação			

Notas sobre as funções do Excel usadas:

(1) Função **RANDBETWEEN(m;n)** ou **ALEATÓRIOENTRE(m;n)** – Função do Excel que devolve um número aleatório inteiro entre m e n, inclusive, com reposição. No nosso caso, em que a população é razoavelmente grande, pois tem dimensão 1513, a probabilidade de obtermos o mesmo número duas

vezes é bastante pequena. No entanto, se isso acontecer, selecionamos outro número para a amostra. Para controlar a existência de valores selecionados iguais, aconselha-se a ordenação dos números gerados.

(2) Função VLOOKUP(i;tabela;j) ou PROCV(i;tabela;j) – Função do Excel que pesquisa, na tabela, o valor da célula i e devolve o valor da tabela que está na mesma linha, mas na coluna j. Por exemplo, a função Vlookup devolveu a palavra “Negro”, da coluna B, que corresponde, na tabela, à etiqueta “125”, da coluna A. Na fórmula, usa-se a referência absoluta (\$A\$3:\$B\$1515) ao definir a tabela para que esta se mantenha quando se replica a fórmula.

(3) Função LEN() ou NÚM.CARACT() – Função do Excel que devolve o número de caracteres inscritos na célula indicada.

(4) Sempre que é executada uma ação no Excel, são recalculados automaticamente os valores gerados aleatoriamente, o que faz alterar a composição da amostra já extraída. Por isso, em cada print aparecem amostras diferentes. Para colmatar esta fragilidade sugerimos usar outra Folha do Excel para analisar as amostras aleatórias que desejarmos. Assim, quando copiamos os dados para uma nova Folha, usamos colar só valores (isto é, sem fórmulas) e estes não voltam a ser recalculados. (É importante fazer **colar especial** e escolher a opção “**colar valores**”).

PROPOSTA DE RESOLUÇÃO

1. Por observação do texto, poderemos arriscar que o valor médio para o número de palavras irá estar entre 3 e 6.

2. Calculou-se a média do número de letras das palavras que compõem a primeira frase do discurso: “Há cem anos, um grande americano, sob cuja sombra simbólica nos encontramos, assinava a Proclamação da Emancipação.”

Assim, obteve-se:

$$\bar{x} = \frac{2 + 3 + 4 + 2 + 6 + 9 + 3 + 4 + 6 + 9 + 3 + 11 + 8 + 1 + 11 + 2 + 11}{17} = 5,6$$

Esta é uma amostra enviesada porque seleciona apenas elementos de um segmento do discurso, pelo que não nos garante uma boa estimativa para a média de todas as palavras da população.

3. Para obtermos uma estimativa do valor médio da população, vamos recolher uma amostra e calcular a sua média. O processo de amostragem deve ser aleatório, para evitar o enviesamento e a falta de representatividade da amostra, que conduziria a que não se obtivesse uma “boa” estimativa. Neste caso, vamos selecionar uma amostra aleatória simples (ver Nota).

Nota: Recordemos os conceitos:

Amostra aleatória e **amostra não aleatória** – Dada uma população, uma amostra aleatória é uma amostra tal que qualquer elemento da população tem alguma probabilidade de ser selecionado para a amostra. Numa amostra não aleatória, alguns elementos da população podem não poder ser selecionados para a amostra.

Amostra aleatória simples – Dada uma população de dimensão N , uma amostra aleatória simples, de dimensão n , é um conjunto de n unidades da população, tal que qualquer outro conjunto de n unidades teria igual probabilidade de ser selecionado. A seleção pode ser feita unidade a unidade, sem reposição, ou selecionando as n unidades todas de uma vez. A propriedade anterior, conduz a que cada elemento da população tenha igual probabilidade de pertencer à amostra.

Para mais informação: https://www.alea.pt/images/cursos/Modulo1-Int_AmostragemFinal.pdf

4. Neste problema, consideramos:

população – conjunto das palavras do discurso

unidades estatísticas – palavras

variável – número de letras da palavra (tamanho da palavra), sendo os *dados* a recolher de tipo *quantitativo discreto*.

4.1. Na folha de Excel chamada **Amostra**, foi inserido o número 10 na célula que recebe o valor da dimensão da amostra. Obtivemos automaticamente 3 amostras de dimensão 10, como se vê na figura seguinte. Para cada uma das amostras existe uma coluna que contém as Etiquetas, uma coluna que contém as Palavras e uma coluna que contém os valores da variável Tamanho (Tam.). Os dados são os que se situam nas colunas relativas ao Tamanho.

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
1	Etiqueta	População	Nº de letras	Dimensão da amostra	Etiqueta	Palavras	Tam.	Etiqueta	Palavras	Tam.	Etiqueta	Palavras	Tam.	Etiqueta	Palavras	Tam.	
2				10	1278	pele	4	568	errados	7	319	a	1				
3	1	ela	2		106	numa	4	1264	juntos	6	418	fatal	5				
4	2	com	3		219	que	3	1152	raparigas	9	446	até	3				
5	3	anos	4		1496	as	2	610	degenere	8	1474	em	2				
6	4	um	2		518	a	1	1416	a	1	202	homens	6				
7	5	grande	6		952	situação	8	1179	as	2	607	uma	3				
8	6	americano	9		805	achar	5	1383	cumes	5	177	Declaração	10				
9	7	sob	3		1507	seja	4	453	liberdade	9	1366	de	2				
10	8	cujas	4		1208	para	4	155	capital	7	206	direitos	8				
11	9	sombra	6		135	na	2	1445	Quando	6	1048	da	2				
12	10	simbólica	9														
13	11	nos	3														
14	12	encontramos	11														
15	13	assinava	8														
16	14	a	1														
17	15	Proclamação	11														
18	16	da	2														

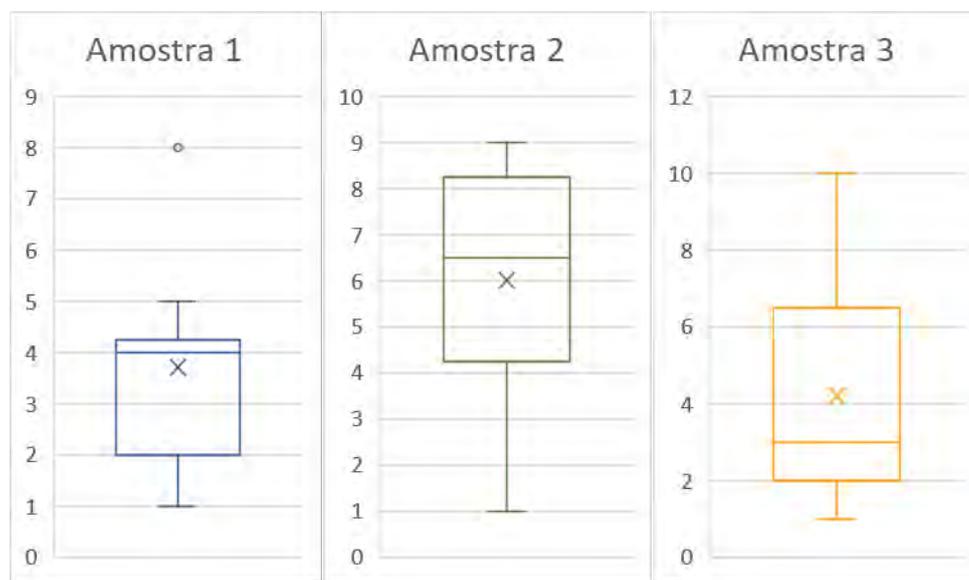
Na folha **Análise** foram colados (usando “colar valores”) os dados das 3 amostras aleatórias obtidas, com o objetivo de proceder à sua análise. Calcularam-se as medidas sugeridas, para cada amostra:

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	AMOSTRA 1		AMOSTRA 2		AMOSTRA 3			AM. 1		AM. 2		AM. 3	
2	x_i		x_i		x_i			Média	3,7	Média	6,0	Média	4,2
3	pele	4	errados	7	a	1		Mediana	4,0	Mediana	6,5	Mediana	3,0
4	numa	4	juntos	6	fatal	5		Mínimo	1	Mínimo	1	Mínimo	1
5	que	3	raparigas	9	até	3		Máximo	8	Máximo	9	Máximo	10
6	as	2	degenere	8	em	2		Amplitude	7	Amplitude	8	Amplitude	9
7	a	1	a	1	homens	6							
8	situação	8	as	2	uma	3							
9	achar	5	cumes	5	Declaração	10							
10	seja	4	liberdade	9	de	2							
11	para	4	capital	7	direitos	8							
12	na	2	Quando	6	da	2							
13													

Podemos agora comparar o valor da média de cada uma das três amostras com o valor da estimativa feita nas questões 1 e 2. Se compararmos com a média obtida com as palavras da primeira frase do discurso, apenas a média da **Amostra 2** se aproxima desse valor.

Relativamente ao intervalo proposto na questão 1, notamos que as médias encontradas para as três amostras estão dentro desse intervalo. Porém, com as três amostras de dimensão 10, continuamos a ter uma grande incerteza sobre o tamanho médio das palavras do discurso. Não é possível dizer qual a que poderá estar mais próxima do valor médio da população.

4.2. Para cada amostra, foi construído, no Excel, o diagrama em caixa de bigodes (ver Nota).

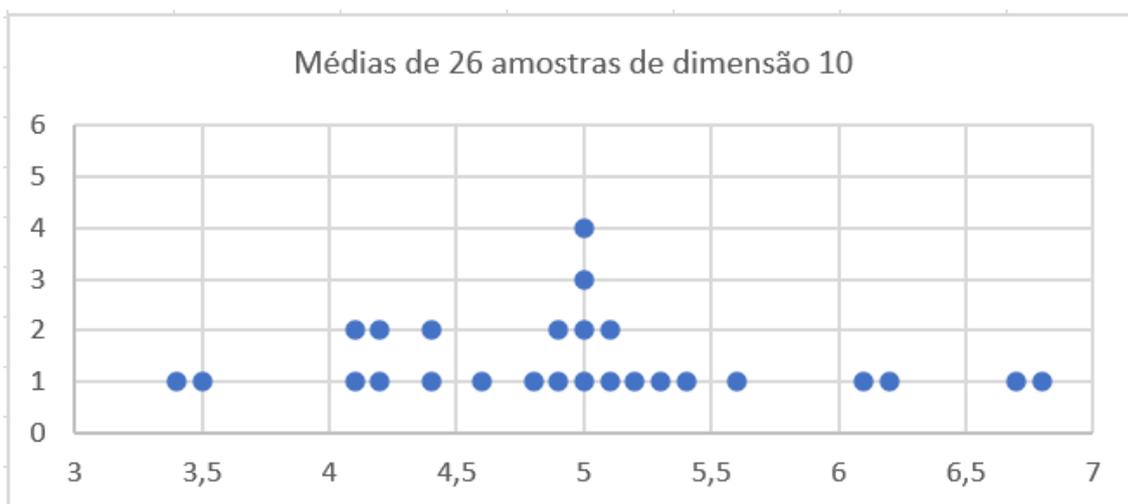


Nota: O diagrama em caixa de bigodes (*box-and-whiskers*) é semelhante ao diagrama de extremos e quartis, mas com a vantagem de representar por pontos os valores considerados *outliers*, no caso de existirem no conjunto de dados em análise. Para mais informação, consultar: https://www.alea.pt/index.php?option=com_content&view=article&id=288&Itemid=1635&lang=pt

Os diagramas evidenciam grande variabilidade para as três amostras selecionadas. Repare-se que a **Amostra 1** apresenta um *outlier*, correspondente a uma palavra de 8 letras. Note-se, no entanto, que a **Amostra 2** tem uma palavra de 9 letras e a **Amostra 3** tem uma palavra de 10 letras, que não são *outliers*. Efetivamente, um dado é *outlier* quando considerado no contexto dos restantes dados da amostra.

No caso da **Amostra 2**, verifica-se um enviesamento para a esquerda, o que está de acordo com o facto de a média ($\bar{x} = 6,0$) ser inferior à mediana ($M_e = 6,5$). Quanto à **Amostra 3**, observa-se um enviesamento para a direita, sendo a média ($\bar{x} = 4,2$) superior à mediana ($M_e = 3,0$), neste caso.

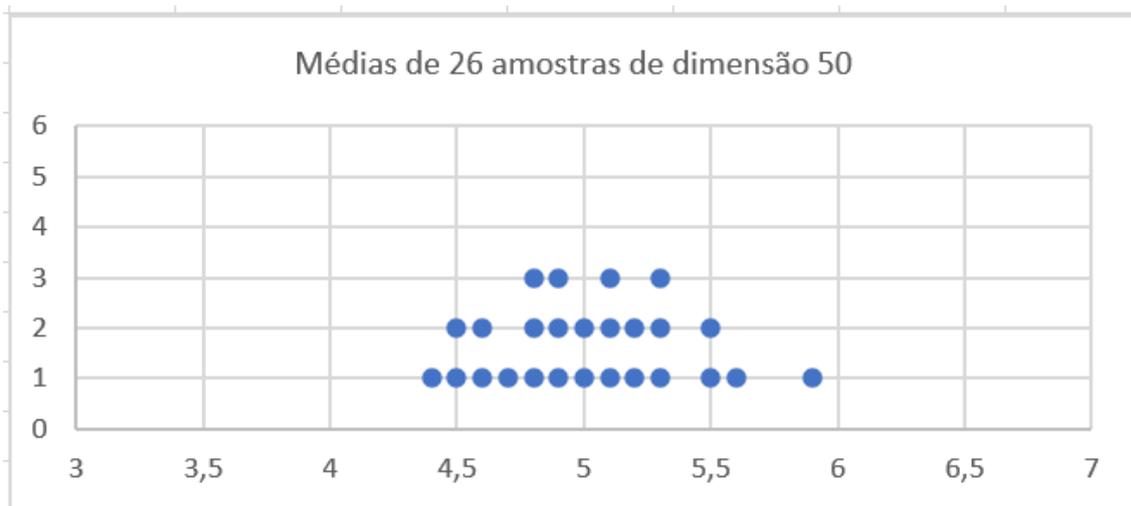
4.3. Apresenta-se o diagrama de pontos que mostra a distribuição das médias de 26 amostras de dimensão 10 que se obtiveram (hipoteticamente, na turma):



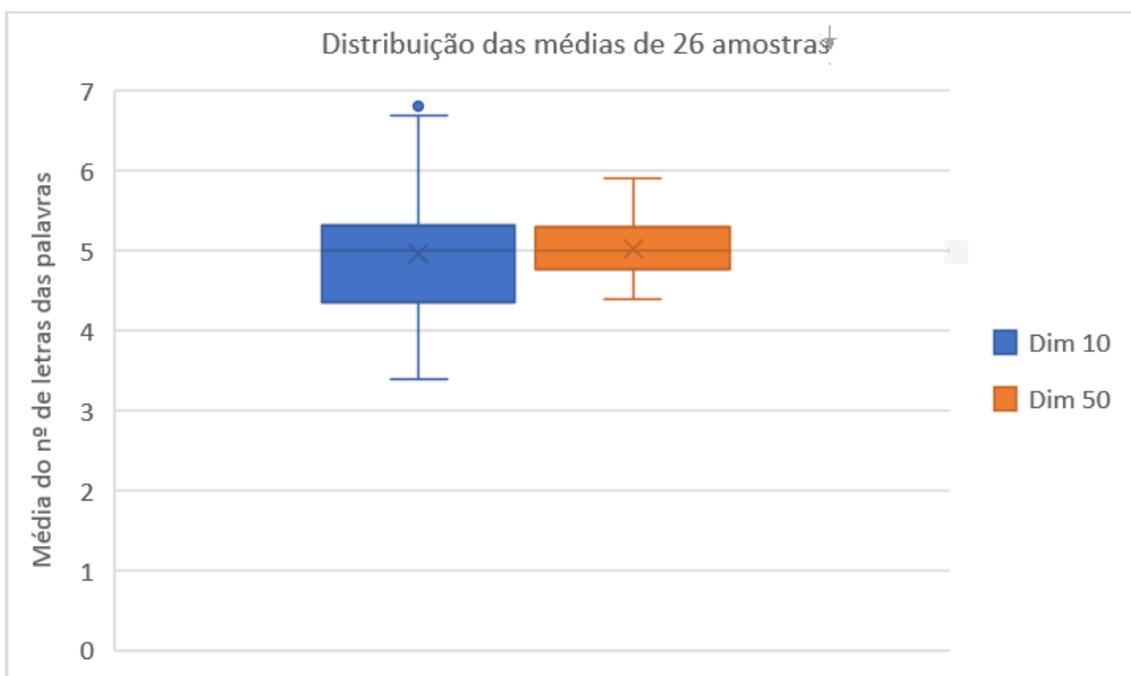
Verifica-se que os valores das médias estão no intervalo $[3,4; 6,8]$, apresentando alguma dispersão, assim como alguma concentração em torno dos valores 4,9 ou 5, o que leva a inferir que o valor médio da população deve estar perto destes valores.

5. Se aumentarmos a dimensão da amostra, esperamos obter estimativas mais próximas da média da população. Quanto maior for a dimensão da amostra, mais informação ela contém sobre a característica populacional que estamos a estudar.

6. Foram recolhidas (na turma) 26 amostras de dimensão 50 e registados os valores das médias das amostras, obtidas da mesma forma que as amostras de dimensão 10. Apresenta-se a distribuição das médias das 26 amostras de dimensão 50 num diagrama de pontos:



Para uma melhor comparação das distribuições das médias de amostras de dimensão 10 e dimensão 50, apresentam-se os diagramas em caixa de bigodes, em paralelo:



Do gráfico anterior podemos concluir que a distribuição das médias das amostras de dimensão 50:

- é aproximadamente simétrica em torno do valor 5;
- os seus valores estão concentrados no intervalo [4.4, 5.9]
- apresenta uma menor dispersão do que a distribuição das médias das amostras de dimensão 10.

7. De acordo com os resultados obtidos na questão anterior, o valor 5 foi considerado uma estimativa razoável para o tamanho médio das palavras do discurso. No Excel, calculou-se o tamanho médio de todas as palavras do discurso e obteve-se o valor 4,9. Assim, confirma-se que a estimativa proposta é adequada.

8. Exemplo de composição (ver Nota)

O nosso objetivo era saber qual é o tamanho médio das palavras do discurso intitulado “I have a dream”, de Martin Luther King, traduzido para língua portuguesa. O discurso tem 1513 palavras. As palavras constituem a população e a variável é o número de letras das palavras. A variável é quantitativa e discreta.

Primeiro, lemos o texto e depois olhámos com atenção para as palavras, tentando ver como eram os diferentes tamanhos. Percebemos que há muitas palavras só com uma letra, por exemplo, “o”, “a”, “e”, mas também há palavras longas e há palavras com 5 letras que aparecem várias vezes, como “vocês” e “sonho”. Para chegarmos a uma aproximação da média da população, selecionámos amostras aleatórias, utilizando o Excel. Cada grupo extraiu várias amostras de dimensão 10 (no total foram recolhidas 26 amostras). Registámos as médias que cada grupo encontrou para as suas amostras e vimos que havia uma grande variação, mas com alguma concentração à volta do 5. Depois, alterámos a dimensão da amostra para 50 e repetimos o processo de obtenção das amostras aleatórias no Excel. Registámos novamente as médias dessas amostras e observámos que estavam menos dispersas e mais concentradas à volta do 5. O gráfico de pontos mostra uma distribuição das médias aproximadamente simétrica (Gráfico 1):

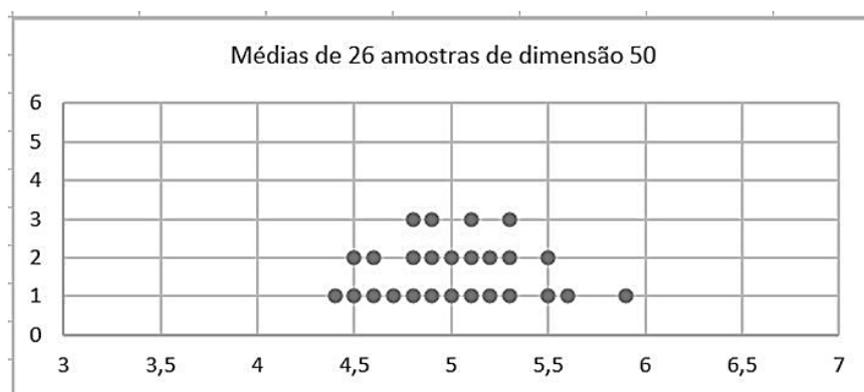


Gráfico 1. Gráfico de pontos coletivo com amostras de dimensão 50

Aumentando a dimensão da amostra (de 10 para 50), podemos dizer, com mais confiança, que 5 letras é uma estimativa aceitável para o tamanho médio das palavras do discurso.

Nota: O texto anterior poderá ser um exemplo de uma composição apresentada pelos alunos. Como proposta de feedback, ou na discussão coletiva final, o(a) professor(a) poderá fazer o seguinte comentário explicativo quanto à precisão de uma estimativa:

Reparem que quando se consideraram as amostras de dimensão 10, 50% das médias dessas amostras estão no intervalo [4,5; 5,5], mas ao passarem para amostras de dimensão 50, mais de 88% dessas médias estão no mesmo intervalo. Como seria de esperar, a dimensão das amostras utilizadas tem influência no processo de estimar o valor médio da população: quanto maior for, mais perto estarão as estimativas umas das outras e esperamos que mais perto da característica da população que estamos a estimar². Porque é que, então, não consideramos amostras de dimensão maior? Porque a recolha de amostras consome tempo e dinheiro, pelo que quando se está a fazer um estudo em que é necessário recolher uma amostra, há que ponderar entre a “precisão” que se pretende e os custos envolvidos!

² No tema de Inferência Estatística estuda-se a forma de, a partir de uma amostra, obter um intervalo que, com determinada **confiança**, normalmente uma confiança de 95%, contenha a característica da população que se pretende estimar. Obtém-se, ainda, uma expressão que relaciona o tamanho da amostra com a amplitude do intervalo.

A EVOLUÇÃO NA VIDA DA MULHER PORTUGUESA NOS ÚLTIMOS 50 ANOS

TAREFA



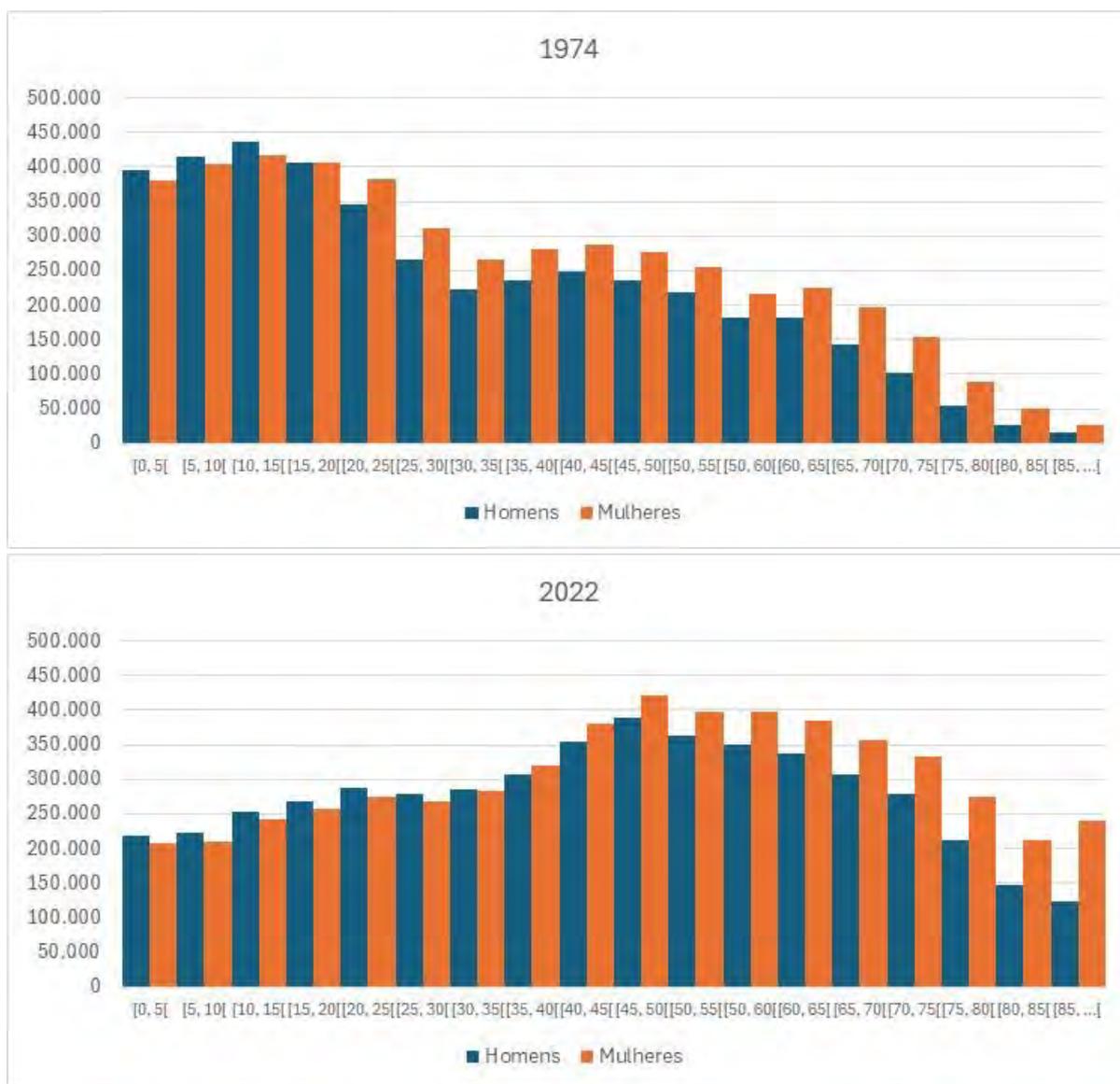
SABIAS QUE...

- ... antes do 25 de Abril de 1974, não havia escolas mistas?
- ... antes do 25 de Abril de 1974, as raparigas não podiam usar saias demasiado curtas, para frequentarem a escola?
- ... antes do 25 de Abril de 1974, as professoras não podiam casar sem autorização especial do Governo?
- ... antes do 25 de Abril de 1974, nem todas as mulheres podiam votar, ao contrário dos homens?
- ... antes do 25 de Abril de 1974, um casal de namorados não se podia beijar na rua?
- ... antes do 25 de Abril de 1974, as mulheres não podiam ser juízas, diplomatas ou polícias?
- ... antes do 25 de Abril de 1974, o casamento válido era o católico e o divórcio não era permitido?
- ... antes do 25 de Abril de 1974, não havia liberdade para ler qualquer livro e ouvir qualquer música?
- ... antes do 25 de Abril de 1974, não era permitido vender Coca-cola?
- ... antes do 25 de Abril de 1974, era necessário ter licença para andar de bicicleta?
- ... antes do 25 de Abril de 1974, a mortalidade infantil era enorme?
- ... poucos anos antes do 25 de Abril de 1974, as mulheres tinham de pedir autorização aos maridos, para viajar para o estrangeiro?
- ... só depois do 25 de Abril de 1974, ficou consagrado na constituição a igualdade de direitos entre homens e mulheres?**

Talvez este preâmbulo te tenha despertado a curiosidade sobre a condição de **ser mulher**, pouco antes do 25 de abril de 1974, que cumpriu agora 50 anos. Vamos analisar alguns dados relativos à população portuguesa, com incidência especial nas mulheres, nomeadamente no que respeita ao emprego, à idade média com que se casavam, à idade média com que tinham o 1.º filho e ao nível de escolaridade mais elevado.

1. Como se distribuía a população em 1974 e 2022?

Os seguintes histogramas representam a distribuição da população residente em Portugal, por sexo e por classes etárias, em 1974 e em 2022³.



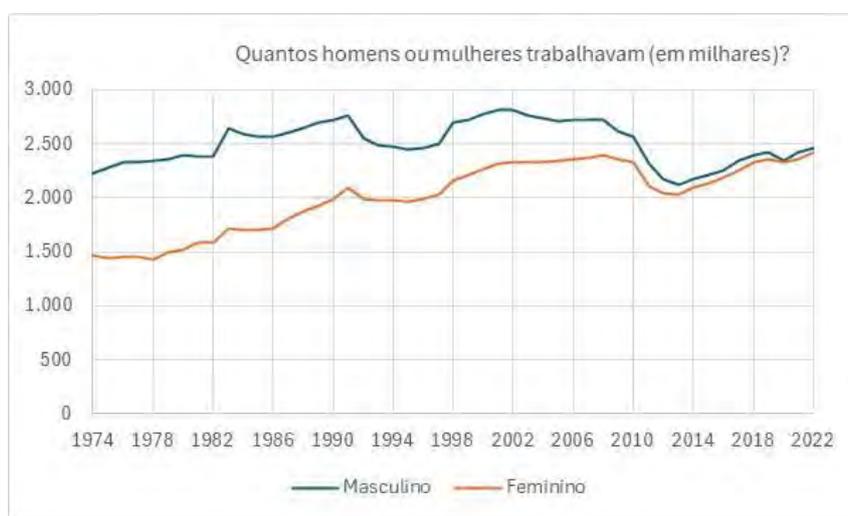
³ Os dados representados nos histogramas foram obtidos em:
<https://www.pordata.pt/portugal/populacao+residente+do+sexo+feminino+total+e+por+grupo+etario-11>
<https://www.pordata.pt/portugal/populacao+residente+do+sexo+masculino+total+e+por+grupo+etario-12>

1.1. Comenta os aspetos que te pareçam mais relevantes, nos gráficos anteriores, destacando as principais características da distribuição da população em cada um dos anos considerados e compara a evolução entre os dois anos em análise.

1.2. Se admitires que a evolução da população nos próximos 50 anos, é idêntica à dos últimos 50 anos, faz um esboço do que esperas para a distribuição da população nessa altura. Justifica.

2. Como se distribuía o emprego, por sexo, nos últimos 50 anos?

Considera o seguinte gráfico⁴ que apresenta a evolução do número de empregados, por sexo, ao longo dos últimos 50 anos⁵.



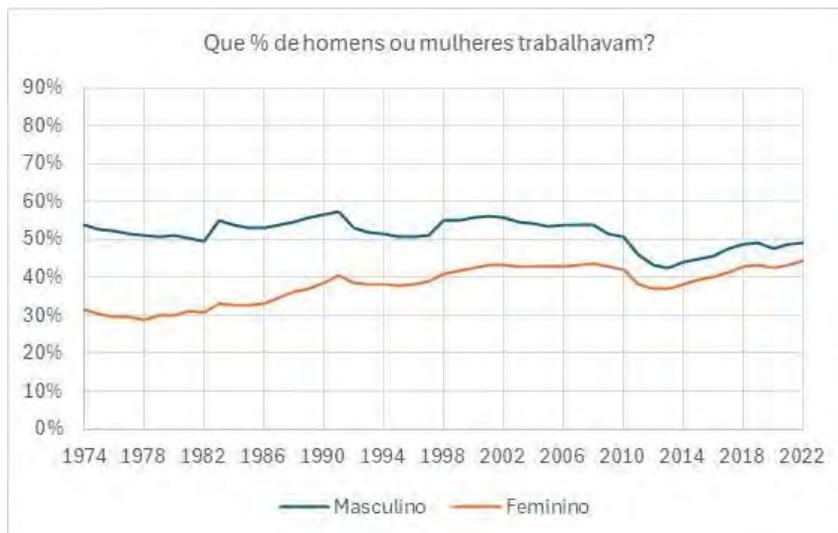
2.1. A partir do gráfico anterior podes concluir que, em cada ano, a percentagem de mulheres empregadas é inferior à percentagem de homens empregados? Justifica.

2.2. Considera agora o seguinte gráfico que pretende transmitir a mesma informação que o gráfico anterior, em que se apresenta, para cada ano, a **percentagem de homens ou mulheres que trabalham**, relativamente ao total de homens ou mulheres residentes, respetivamente, nesse ano:

⁴ Os dados a que se refere o gráfico estão disponíveis em:

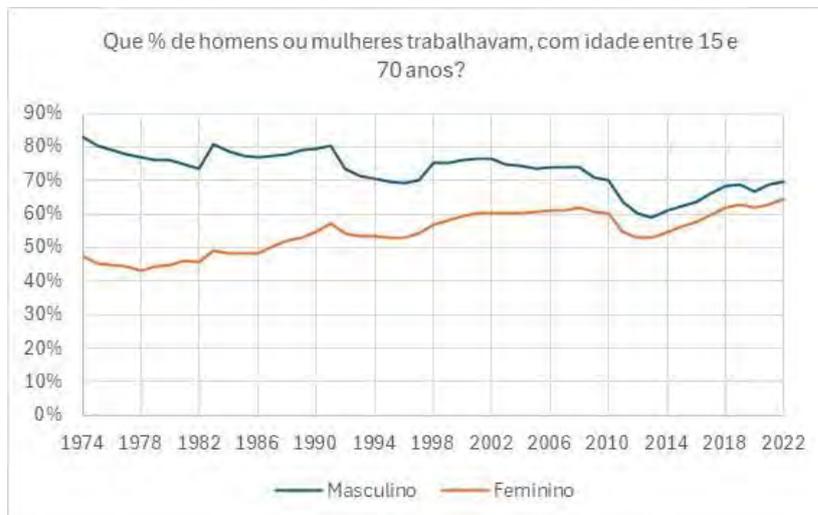
<https://www.pordata.pt/portugal/populacao+empregada+total+e+por+sexo-30>

⁵ Em 1983, 1992, 1998 e 2011, houve alterações para a classificação da situação de empregado e de desempregado, que são refletidas nos gráficos de linhas. Para mais detalhes, consultar a Metainformação na página da PORDATA referida na nota anterior.



- i) Comparando os dois gráficos anteriores, há alguma(s) característica(s) que te mereça(m) destaque?
- ii) Qual dos dois gráficos consideras que representa melhor a informação sobre o emprego dos homens e mulheres?

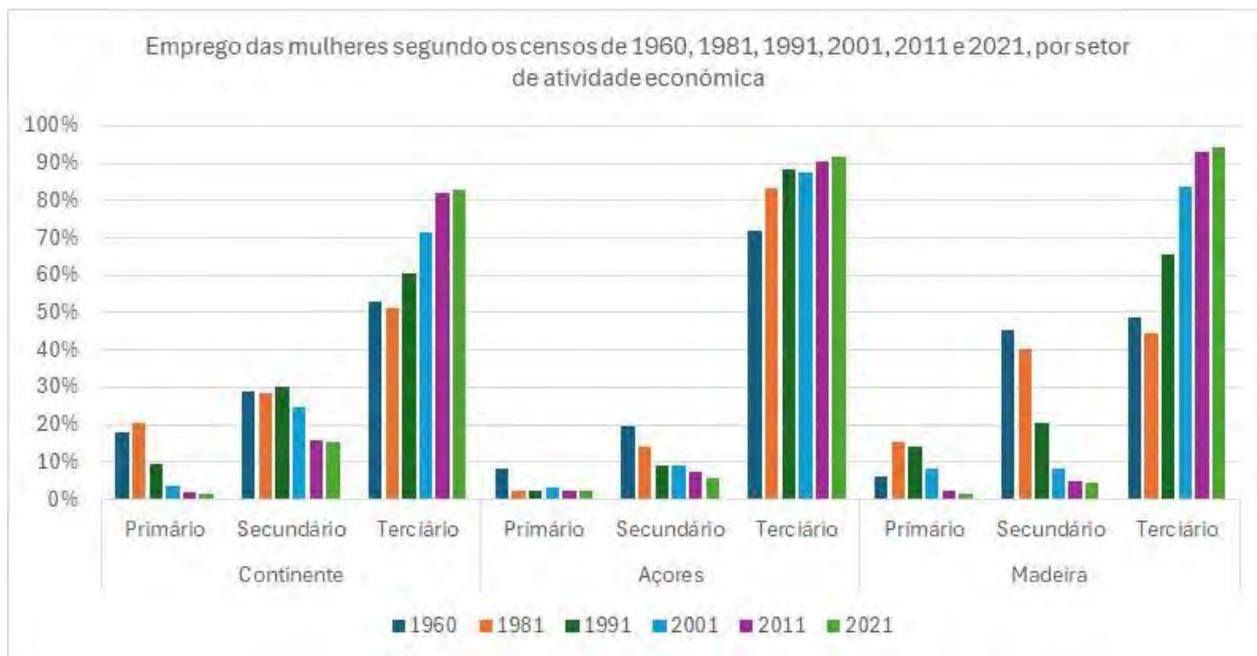
2.3. Considera ainda, sobre a mesma situação em estudo, o seguinte gráfico que apresenta a **percentagem de homens e mulheres que trabalham**, mas agora relativamente ao total de residentes com idade igual ou superior a 15 anos e inferior a 70 anos:



Se pretendesses transmitir a informação sobre a situação da população empregada em Portugal, nos últimos 50 anos, qual dos gráficos escolherias? Justifica a tua resposta.

3. Como se distribuía o emprego das mulheres, nos últimos 6 censos, por setor de atividade económica⁶?

Considera a seguinte representação gráfica⁷, onde é apresentada a forma como se distribuía o emprego das mulheres nos últimos 6 censos, por setor da atividade económica, e em que se consideram separadamente o Continente, a Região Autónoma dos Açores e a Região Autónoma da Madeira:



3.1. A partir do gráfico anterior, responde às seguintes questões:

- Em qual das regiões (Continente, Açores ou Madeira) era superior a 70% o emprego no setor terciário, segundo o censo de 1960?
- Segundo o censo de 1981, quais os valores, aproximados, para a percentagem de mulheres com emprego no Continente, nos diferentes setores de atividade económica?
- Em qual das regiões se manteve sempre superior a 15%, a percentagem de mulheres com emprego no setor secundário?

⁶ A atividade económica inclui a agricultura, a indústria transformadora, a construção e obras públicas, o comércio e outros ramos de atividade em que se podem agrupar quem produz o mesmo tipo de bens e de serviços. É frequente agrupar as atividades económicas em três grandes setores: 1. Primário, incluindo agricultura, floresta, caça, pesca e extração mineral; 2. Secundário, incluindo indústria transformadora e construção; e 3. Terciário, incluindo os serviços, tais como comércio, transportes, administração pública, educação ou saúde.

⁷ Os dados a que se refere o gráfico estão disponíveis em:
<https://www.pordata.pt/municipios/populacao+empregada+do+sexo+feminino+segundo+os+censos+tal+e+por+setor+de+atividade+economica-158>

3.2. Comenta alguns aspetos que consideres relevantes sobre a informação transmitida pelo gráfico anterior.

4. Qual a idade média ao primeiro casamento das mulheres, nos últimos 50 anos?

A seguinte tabela⁸ apresenta a idade média ao primeiro casamento, por sexo:

Anos	Sexo	
	Masculino	Feminino
1974	26,4	24,1
1975	25,7	23,6
1976	25,5	23,6
1977	25,7	23,7
1978	25,7	23,7
1979	25,4	23,3
1980	25,4	23,3
1981	25,4	23,3
1982	25,3	23,2
1983	25,2	23,2
1984	25,4	23,4
1985	25,6	23,6
1986	25,7	23,7
1987	25,8	23,8
1988	25,9	23,9
1989	26,1	24,1
1990	26,2	24,2
1991	26,3	24,4
1992	26,4	24,5
1993	26,5	24,7
1994	26,7	24,8
1995	26,8	24,9
1996	27,0	25,1
1997	27,1	25,3
1998	27,2	25,4
1999	27,3	25,6
2000	27,5	25,7
2001	27,8	26,1
2002	28,0	26,4

⁸ Os dados a que se refere a tabela estão disponíveis em:

<https://www.pordata.pt/portugal/idade+media+ao+primeiro+casamento++por+sexo-421>

2003	28,4	26,8
2004	28,6	27,0
2005	28,9	27,3
2006	29,1	27,5
2007	29,4	27,8
2008	29,7	28,1
2009	30,2	28,6
2010	30,8	29,2
2011	31,1	29,5
2012	31,4	29,9
2013	31,7	30,2
2014	32,1	30,6
2015	32,5	31,0
2016	32,8	31,3
2017	33,2	31,6
2018	33,6	32,1
2019	33,9	32,4
2020	34,9	33,4
2021	34,3	32,9
2022	35,1	33,7

4.1. Considera os dados apresentados na tabela anterior e mostra, graficamente, como evoluiu nos últimos 50 anos a idade média da mulher ao primeiro casamento.

4.2. Encontras alguma razão para os resultados obtidos na alínea anterior?

4.3. A idade média dos homens ao primeiro casamento é idêntica à das mulheres? Justifica a tua resposta.

5. Qual a idade média da mãe ao nascimento do primeiro filho, nos últimos 50 anos? Considera o seguinte gráfico⁹ que apresenta como evoluiu, nos últimos 50 anos, a idade média da mãe ao nascimento do primeiro filho:

⁹ Os dados a que se refere a tabela estão disponíveis em:

<https://www.pordata.pt/portugal/idade+media+da+mae+ao+nascimento+do+primeiro+filho-805>



5.1. Pensas que a representação gráfica anterior é a mais adequada para transmitir a informação que se pretende?

5.2. Que outro tipo de representação gráfica considerarias mais adequada, caso tenhas respondido que não?

5.3. Comenta a informação transmitida pelo gráfico anterior.

6. Considera a seguinte tabela que apresenta a população (em milhares) residente em Portugal do sexo feminino com idade entre 16 e 89 anos, total e por nível de escolaridade completo mais elevado, de 1998 a 2023.

Anos	Total	Nível de escolaridade					
		Sem nível de escolaridade	Básico - 1º ciclo	Básico - 2º ciclo	Básico - 3º ciclo	Secundário e pós-secundário	Superior
1998	4.431,6	1.078,7	1.424,9	618,6	549,9	454,4	305,0
1999	4.464,4	1.080,8	1.401,7	627,1	564,4	477,8	312,6
2000	4.507,0	1.042,8	1.411,6	628,5	590,8	510,0	323,2
2001	4.546,5	1.023,7	1.401,8	634,4	606,7	529,9	350,0
2002	4.576,5	978,2	1.424,2	640,0	625,1	530,6	378,5
2003	4.601,0	935,8	1.401,6	633,6	634,6	563,5	432,0
2004	4.617,9	860,4	1.377,6	629,7	665,3	593,3	491,7
2005	4.632,8	832,1	1.353,8	633,9	684,7	625,9	502,5
2006	4.652,4	792,4	1.339,4	644,1	702,7	645,4	528,4
2007	4.673,5	765,6	1.347,3	652,8	720,5	635,5	551,8
2008	4.693,8	737,3	1.343,8	603,9	798,0	629,1	581,7
2009	4.710,5	681,9	1.344,7	577,2	841,2	665,5	600,0
2010	4.730,1	649,3	1.312,2	573,4	835,3	717,0	642,8

2011	4.650,2	653,2	1.159,7	496,9	860,3	773,1	707,0
2012	4.640,8	615,2	1.140,1	453,2	841,5	829,7	761,0
2013	4.627,4	571,7	1.124,5	436,6	823,6	853,6	817,4
2014	4.627,6	533,0	1.092,6	424,8	804,2	876,6	896,3
2015	4.613,4	498,2	1.061,3	423,1	816,3	890,5	924,1
2016	4.614,5	466,4	1.035,7	414,0	830,1	910,0	958,3
2017	4.615,1	435,0	1.035,6	411,8	796,4	948,4	987,9
2018	4.608,4	401,4	1.027,2	392,3	778,5	988,2	1.020,7
2019	4.629,4	380,8	998,0	376,8	798,0	1.025,4	1.050,5
2020	4.651,8	336,2	988,3	376,5	799,8	1.032,8	1.118,1
2021	4.628,9	294,8	980,1	354,9	778,8	1.053,7	1.166,7
2022	4.635,0	258,1	945,9	343,7	778,7	1.114,1	1.194,4
2023	4.644,5	227,8	920,0	354,0	777,3	1.161,0	1.204,4

Decide como fazer um resumo dos dados e prepara uma apresentação oral da informação contida na tabela. Talvez já tenhas ouvido dizer que “um gráfico vale mais do que mil palavras”, desde que o gráfico seja adequado... Selecciona, então, gráficos que julgues adequados para fazer essa apresentação aos teus colegas e professor(a) e explica as principais conclusões a que chegaste.

NOTAS PARA O PROFESSOR

A exploração da tarefa **A evolução na vida da mulher portuguesa nos últimos 50 anos** é adequada às Aprendizagens Essenciais de Matemática A, Matemática B, MACS e Matemática dos Cursos Profissionais.

No ano em que se comemoram os 50 anos do 25 de abril, pretende-se com esta tarefa dar a conhecer aos jovens alguns direitos negados às mulheres antes do 25 de abril e recorrer aos conhecimentos de estatística para melhor compreender problemas sociais.

Nesta tarefa pretende-se analisar alguns aspetos da evolução na vida das mulheres e da população portuguesa nos últimos 50 anos, de forma a contribuir para consciencializar os alunos das mudanças que ocorreram na sociedade portuguesa com o 25 de abril de 1974. Esta tarefa constitui ainda uma oportunidade para o trabalho interdisciplinar, o desenvolvimento da comunicação matemática e do espírito crítico.

A resolução da tarefa, a partir da apresentação de diversas representações gráficas de dados, tais como histograma, gráfico de linhas, diagrama de barras pretende levar os alunos a ler, interpretar e analisar estudos estatísticos e reconhecer que algumas representações são mais adequadas do que outras.

Tema: Estatística

Tópicos e subtópicos:

Problema estatístico

- Variabilidade

População, amostra e variável

Dados univariados

- Dados contínuos

- Histograma

- Medidas de localização

- Medidas de dispersão

- Propriedades das medidas

Dados bivariados

- Dados quantitativos

- Gráfico de linhas

Objetivos de aprendizagem:

- Reconhecer o papel relevante desempenhado pela Estatística em todos os campos do conhecimento.
- Conhecer e interpretar situações do mundo que nos rodeia em que a variabilidade está presente.

- Selecionar representações gráficas adequadas para cada tipo de dados identificando vantagens/inconvenientes.
- Reconhecer que algumas representações gráficas são mais adequadas que outras para comparar conjuntos de dados.
- Entender que um gráfico de linhas é um caso particular de um diagrama de dispersão, em que se pretende estudar a evolução de uma das variáveis relativamente a outra variável, de um modo geral o tempo, e em que se unem, por linhas, os pontos representados.

Esta tarefa permite recordar as diferentes representações gráficas estudadas no ensino básico, incentivar a análise crítica dessas representações e promover a comunicação matemática.

Duração prevista: Duas aulas de 50 minutos.

Materiais e recursos:

- Calculadora gráfica, computador.
- Internet, folha de cálculo.

Estratégias de implementação:

Esta tarefa apresenta diversas representações gráficas de dados sobre um tema muito relevante da história contemporânea, em particular, sobre a evolução na vida das mulheres nos últimos 50 anos. Pretende-se evidenciar a importância da estatística para o conhecimento das mudanças sociais, em particular, na área dos direitos das mulheres.

Recomendamos que a tarefa seja resolvida em grupos, para facilitar a análise dos dados disponibilizados e promover a discussão entre pares. Esta tarefa permite estabelecer ligações a outras áreas do conhecimento pelo que os alunos poderão investigar, aprofundar e discutir o tema em outras áreas disciplinares.

No final, poderá ser promovida uma discussão coletiva sobre o tema estudado, pedindo a um ou mais grupos que apresentem as suas conclusões. A partir das respostas apresentadas pelos diversos grupos, deverá ser feita, de forma colaborativa, uma síntese das principais ideias associadas a cada uma das representações, das suas potencialidades e adequação às diferentes situações. Esta tarefa é potenciadora da comunicação matemática.

No final da aula, os alunos devem refletir individualmente sobre o seu trabalho e o que aprenderam. Tal pode ser concretizado através de um pequeno questionário.

Dificuldades esperadas e ações do professor:

Os alunos podem revelar algumas dificuldades, nomeadamente em:

- Analisar e compreender os dados disponibilizados através de diferentes representações gráficas e em trabalhar com tanta quantidade de informação.
- Compreender a adequação das diferentes representações às situações apresentadas.

O professor, deverá acompanhar o trabalho dos diferentes grupos, solicitando aos mesmos que verbalizem a interpretação dos dados, dando sugestões ou colocando questões que contribuam para o desenvolvimento do trabalho. Deverá ainda solicitar aos alunos que apresentem as respostas por escrito justificando as suas respostas.

Avaliação:

A avaliação poderá ter um carácter formativo. Durante a aula o professor poderá dar aos alunos feedback oral sobre o seu desempenho na resolução da tarefa.

PROPOSTA DE RESOLUÇÃO

1.1. Relativamente ao ano de 1974, os aspetos que se poderão considerar mais relevantes são os seguintes:

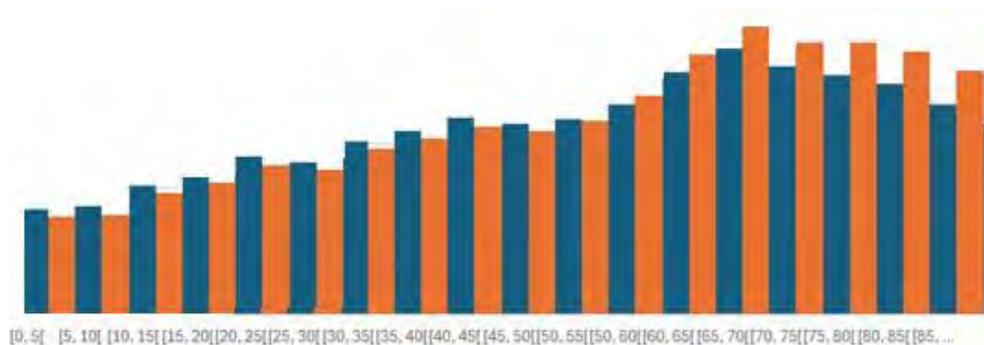
- A população é acentuadamente jovem, dando a entender que a taxa de natalidade era elevada, principalmente nos anos anteriores a 1974, mas também era elevada a taxa de mortalidade, sendo esta superior nos homens;
- Nas classes etárias até aos 15 anos, predominam os elementos do sexo masculino, o que vai de encontro ao facto de nascerem mais rapazes do que raparigas¹⁰;
- Na classe etária dos 15 aos 20, o número de elementos do sexo masculino é aproximadamente igual ao número de elementos do sexo feminino, o que indicia que a taxa de mortalidade até aos 15 anos, é mais elevada nos rapazes, do que nas raparigas;
- Verifica-se um decréscimo acentuado dos homens relativamente às mulheres entre os 20 e os 30 anos, o que poderá estar relacionado com a guerra do ultramar.
- A partir dos 25 anos as mulheres predominam relativamente aos homens, acentuando-se esta diferença nas últimas classes etárias.

Relativamente ao ano de 2022, os aspetos que se poderão considerar mais relevantes são os seguintes:

- Ao contrário do que se verificava em 1974, a população é acentuadamente de meia-idade;
- Entre 1974 e 2022, verificou-se uma acentuada diminuição da taxa de natalidade, pelo que a população jovem diminuiu relativamente à restante;
- O facto de ter aumentado fortemente a população com idade superior a 65 anos, é sinal do aumento significativo da esperança de vida, sobretudo das mulheres;

Quando se compara a distribuição da população em 2022, relativamente a 1974, verifica-se um acentuado envelhecimento, sintoma do decréscimo da taxa de natalidade e do aumento da esperança de vida, nos últimos 50 anos. O facto de a mulher começar a querer ter uma carreira profissional, a par dos homens, conduziu a que já não tivesse tanta disponibilidade para ter muitos filhos. Por outro lado, a melhoria das condições de vida, nomeadamente com o Serviço Nacional de Saúde, faz com que as pessoas tenham uma maior esperança de vida.

1.2. Admitindo que a evolução da população nos próximos 50 anos é idêntica à dos últimos 50, espera-se uma diminuição do número de jovens e um aumento dos idosos. Assim, o pico da distribuição terá tendência a deslocar-se para a direita, para a classe etária dos mais velhos, tal como se esboça a seguir:



¹⁰ Veja-se a TarefALEA 7 - O Envelhecimento da população da UE 27 (UE - 27), com destaque para a população portuguesa. Disponível em:
https://alea.ine.pt/index.php?option=com_content&view=article&id=972&Itemid=2166&lang=pt

2.1. Não se pode tirar essa conclusão, porque o que o gráfico apresenta são as frequências absolutas com que determinada característica se verifica em duas populações de dimensão diferente, pelo que só tem sentido utilizar as frequências relativas.

2.2.

i) Comparando os dois gráficos anteriores, verifica-se que a informação transmitida pelo primeiro, no sentido de que a partir de 2010, havia praticamente tantos homens como mulheres, no mercado de trabalho, não é correta. Quando se considera o segundo gráfico, com as percentagens, verifica-se que, embora a percentagem de mulheres a trabalhar tenha vindo a aumentar, há sempre uma diferença, relativamente aos homens de, pelo menos, cerca de 5%.

De realçar ainda, nos dois gráficos, algumas quedas no seguimento natural dos dados, nomeadamente, em 1983, 1992 e 2011. Esses factos estão relacionados com a alteração do que se considera empregado. Por exemplo, em 2011, *a idade da população ativa e empregada passou de 15 e mais anos para 16 aos 89 anos*¹¹.

ii) Pelo que ficou dito anteriormente, a escolha recai sobre o gráfico que apresenta as percentagens.

2.3. Seria escolhido este último gráfico, pois traduz de forma mais eficaz a realidade, já que apresenta a percentagem dos homens ou mulheres que trabalham, relativamente àqueles que estão em condições etárias para trabalhar.

3.

3.1.

i) Na região dos Açores.

ii) Aproximadamente 20%, 30% e 50%, nos setores primário, secundário e terciário, respetivamente.

iii) No Continente.

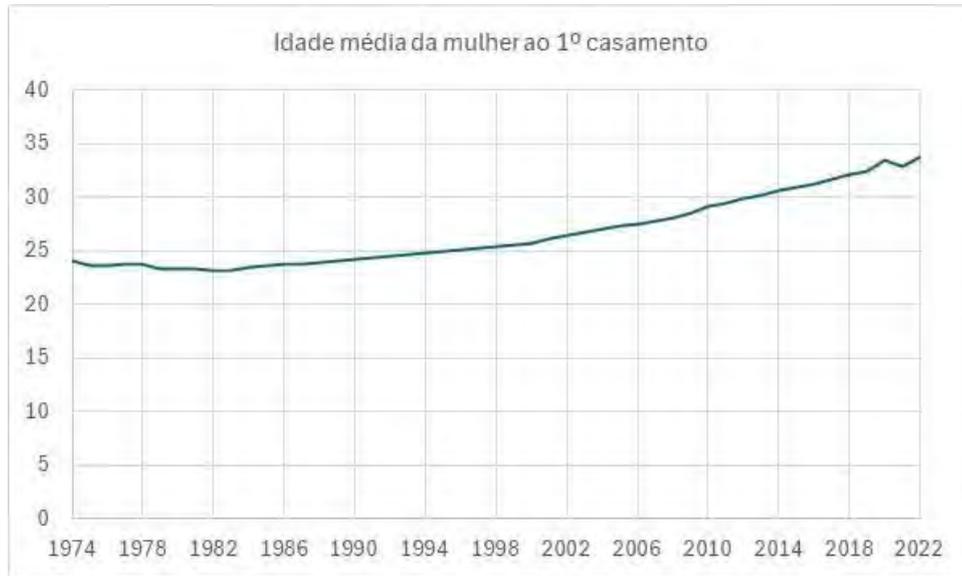
3.2. Alguns aspetos a realçar, são:

- Como seria de esperar, o setor dominante é o setor terciário e o menos representado é o primário. Recorda-se que no setor primário estão incluídas a agricultura, pecuária, silvicultura, extração mineira e pescas, que são serviços com prevalência do sexo masculino.
- Um dos aspetos que merece maior realce, é o facto de na Região Autónoma dos Açores, o trabalho dominante das mulheres ser sempre no setor terciário e ser sempre superior a 70%, ao contrário das outras regiões.
- É interessante verificar que os censos de 1960 e 1981, revelam que na Região Autónoma da Madeira, a percentagem de mulheres a trabalhar nos setores secundário e terciário é aproximadamente igual.
- Um aspeto também a destacar é o facto de o setor primário quase não ter expressão na Região Autónoma dos Açores.

¹¹ <https://www.pordata.pt/portugal/populacao+empregada+total+e+por+sexo-30>

4.

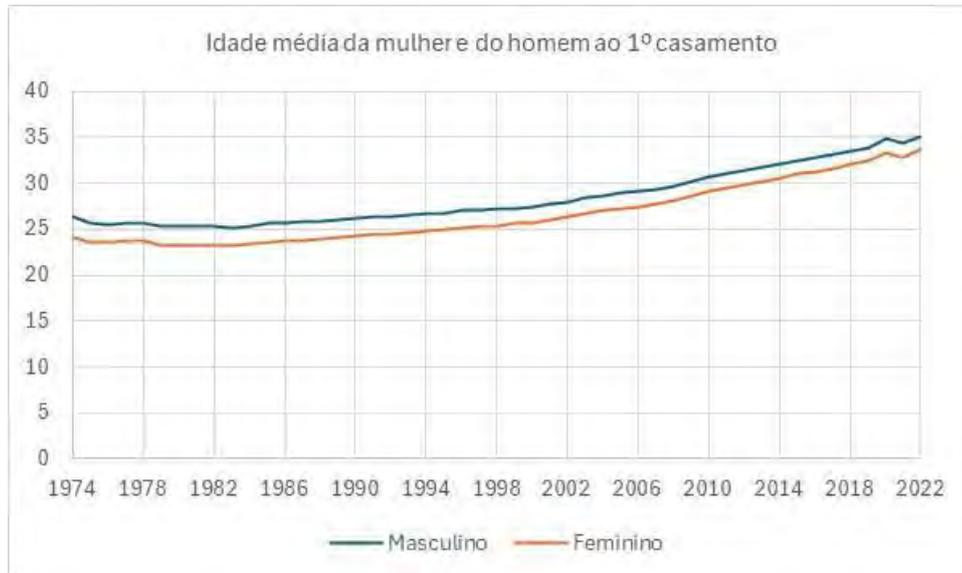
4.1. Consideramos o seguinte gráfico de linhas que é o mais adequado para mostrar a forma como evoluiu a idade média da mulher ao 1º casamento:



Como se verifica, a idade média da mulher ao 1º casamento, que nos anos 70 andava à volta dos 24 anos, aumentou cerca de 10 anos, situando-se perto dos 34 anos em 2022.

4.2. Uma razão possível poderá ser uma maior preocupação com a carreira profissional, que relegou para segundo plano o casamento.

4.3. A partir dos dados da tabela, poder-se-ia fazer a diferença entre os dados relativos ao homem e os dados relativos à mulher, ou uma segunda opção, que aqui foi seguida, seria representar os dois conjuntos de dados, como se apresenta a seguir:



Como se verifica do gráfico anterior, a idade média do homem ao 1º casamento é sistematicamente superior à da mulher, em cerca de 2 anos.

5.

5.1. A representação gráfica utiliza um gráfico com barras (**Nota:** não é um gráfico **de** barras, pois este é utilizado para representar dados qualitativos ou quantitativos discretos, agrupados em classes, numa tabela de frequências, quer absolutas quer relativas). Aqui, temos um gráfico com barras, que está a representar os dados, ou seja, indica para cada ano, a idade média da mãe ao 1º filho. Não parece ser uma representação adequada, já que a informação relevante que se pretende transmitir é a evolução ao longo dos anos da idade média da mulher ao 1º filho, que se poderia obter unindo os pontos superiores das barras e eliminando as barras para ficar mais claro.

5.2. Da resposta anterior, deduz-se que a representação mais adequada seria o gráfico de linhas.

5.3. Do gráfico, concluímos que a idade média da mulher ao 1º filho, que rondava os 24 anos, nos anos 70, aumentou para cerca de 31 anos, em 2022. O facto de a idade da mulher ao nascimento do 1º filho estar perto de 31 anos também contribui para o envelhecimento da população, já que agora não tem tantos filhos como na altura em que a idade do 1º filho rondava os 24 anos.

6. Resposta livre.

DOÇURA OU TRAVESSURA (E A QUESTÃO DA DESIGUALDADE)

TAREFA



PARTE I

Por altura do Halloween, 10 alunos do 1º Ciclo saíram da escola e foram bater às portas, na conhecida brincadeira “Doçura ou Travessura”. No final, contaram o total de rebuçados que tinham conseguido ganhar, que foi 122 rebuçados.

Abaixo, mostra-se uma tabela com o número de rebuçados que cada aluno ganhou, estando a tabela ordenada da criança que ganhou menos para a que ganhou mais:

Alunos	Nº de rebuçados
Marta	3
Joana	5
Diogo	7
Soraia	8
Fábio	9
Beatriz	11
Tiago	12
Alice	15
Igor	19
Pedro	33

MENOS FAVORECIDO

MAIS FAVORECIDO

1. O que mais te chama a atenção no conjunto dos dados? Explica a tua resposta.
2. Qual foi a média do número de rebuçados obtidos pelos 10 alunos? A média dá-te, neste caso, uma informação útil acerca do número de rebuçados que os vários

alunos ganharam? (Sugestão: Recorrendo à tecnologia, constrói um gráfico com barras ou um pictograma que represente os dados e onde se assinala o valor da média. O que podes concluir?)

3. Parece-te que a distribuição do número de rebufados recebidos é simétrica ou assimétrica? Justifica a resposta a partir da construção de um diagrama de caule-e-folhas com 4 caules.

4. Calcula a mediana do número de rebufados recebidos pelos alunos. Como se situam a média e a mediana, uma em relação à outra? Como interpretar esse facto?

5. Determina os quartis da distribuição de rebufados. Explica o significado de cada um dos valores obtidos.

6. Preenche uma nova coluna na tabela que mostre a **parcela** do total de rebufados, em percentagem, recebida por cada aluno. Usa aproximação às centésimas.

Alunos	Nº de rebufados	Parcela do total (%)
Marta	3	
Joana	5	
Diogo	7	
Soraia	8	
Fábio	9	
Beatriz	11	
Tiago	12	
Alice	15	
Igor	19	
Pedro	33	

7. Repara que o grupo de 10 alunos está automaticamente dividido em 10 estratos ordenados, do menos favorecido para o mais favorecido, sendo cada estrato composto por 1 aluno, ou seja, tendo 10% dos elementos do grupo. Assim, considera as questões:

7.1. Será verdadeira a afirmação: “Os **30% mais favorecidos** conseguiram angariar mais de metade do total de rebufados”? Justifica.

7.2. Quantas vezes é maior a quantidade de rebufados dos **10% mais favorecidos** do que a quantidade de rebufados dos **10% menos favorecidos**?

7.3. Quantas vezes é maior a quantidade de rebufados dos **20% mais favorecidos** do que a quantidade de rebufados dos **20% menos favorecidos**?

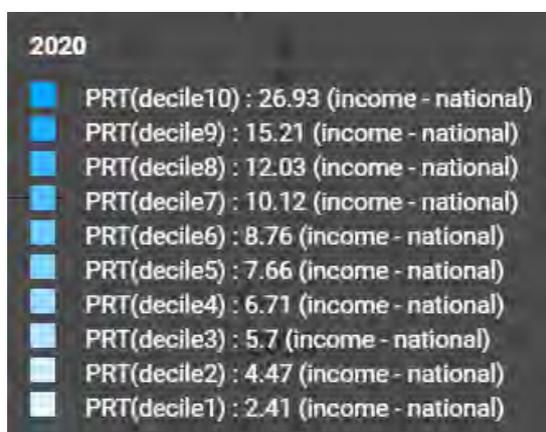
PARTE II

Nas estatísticas de distribuição de rendimentos de uma população, os indivíduos são ordenados por ordem crescente do seu rendimento. Depois divide-se a população em 10 estratos de igual tamanho. Cada estrato (ou decil) contém 10% da população. O primeiro estrato corresponde aos 10% de pessoas com rendimentos mais baixos e o último estrato corresponde aos 10% de pessoas com rendimentos mais elevados. Para cada estrato é calculada a **parcela** (em percentagem) do rendimento global obtida por esse conjunto de pessoas.

Segundo dados do Banco Mundial, relativos a 2020 (ver nota abaixo), a tabela seguinte mostra a parcela de rendimento (em percentagem) que ganhou cada estrato de 10% da população portuguesa. Por exemplo, os **10% mais pobres** ganharam 2,41% do rendimento total e os **10% mais ricos** ganharam 26,93% do rendimento total.

Estratos da população	Parcela do rendimento total (%)
1	2,41
2	4,47
3	5,70
4	6,71
5	7,66
6	8,76
7	10,12
8	12,03
9	15,21
10	26,93

Como podes ver, as parcelas de rendimento recebidas por cada estrato da população portuguesa equivalem (grosso modo) às parcelas de rebuçados que os 10 alunos ganharam no Halloween. Com essa informação, faz uma análise estatística da desigualdade de distribuição de rendimentos em Portugal, referente ao ano de 2020.



Nota: Dados obtidos do World Bank
(<https://pip.worldbank.org/poverty-calculator>)

World Bank (2023). Poverty and Inequality Platform (version 20230919_2017_01_02_PROD) [data set]. pip.worldbank.org. Acedido em 2023-10-20.

NOTAS PARA O PROFESSOR

A exploração da tarefa **Doçura ou travessura** é adequada às Aprendizagens Essenciais de Matemática A, Matemática B, MACS e Matemática dos Cursos Profissionais.

A tarefa tem um grande foco no desenvolvimento da Educação para a Cidadania e numa perspetiva de analisar as desigualdades de distribuição de rendimento, no nosso país, promovendo a autonomia dos alunos, o trabalho interdisciplinar, a comunicação matemática e o desenvolvimento do espírito crítico.

Tema: Estatística

Tópicos e subtópicos:

Dados univariados

- Organização de dados
- Medidas de localização
- Propriedades das medidas

Objetivos de aprendizagem:

- Selecionar representações gráficas adequadas para cada tipo de dados identificando vantagens/inconvenientes, relembrando a construção de gráficos de barras, diagramas de caule-e-folhas e diagramas de extremos-e-quartis.
- Interpretar as medidas de localização: média (\bar{x}), mediana (Me), moda(s) (Mo) e percentis (quartis como caso especial) na caracterização da distribuição dos dados, relacionando-as com as representações gráficas obtidas.
- Reconhecer que existem situações em que é preferível utilizar, como medida de localização do centro da distribuição dos dados, a mediana em vez da média, e como medida de dispersão a amplitude interquartil em vez do desvio padrão, apresentando exemplos simples.

Duração prevista: Duas aulas de 50 minutos.

Materiais e recursos:

- Calculadora gráfica, computador.
- Internet, folha de cálculo.

Estratégias de implementação

A primeira parte da tarefa, pretende relembrar conceitos lecionados no ensino básico, referentes à organização de dados e construção de um diagrama de caule-e-folhas, mas simultaneamente, pôr os alunos a refletir sobre uma hipotética situação de distribuição de doces no Halloween e sobre o facto de esta não ser equitativa para todas as crianças.

Esta reflexão, guiada pelas questões da tarefa, conduz os alunos a ponderarem sobre as propriedades das medidas, em particular o uso da média e da mediana.

Recomendamos que a tarefa possa ser resolvida em pares, para promover a discussão sobre a desigualdade na distribuição dos doces, que servirá como mote para a realização da segunda parte da tarefa.

A segunda parte da tarefa é mais aberta, levando os alunos a uma discussão sobre a desigualdade de rendimentos em Portugal. É natural que alguns dos alunos sejam levados a verificar como será essa distribuição noutros países.

No final, poderá ser promovida uma discussão coletiva sobre o tema da desigualdade de rendimentos, pedindo aos alunos que mostrem como organizaram os dados, que tipos de representações gráficas utilizaram na resolução e como justificam a sua escolha. Finalmente, a partir das apresentações realizadas e das aprendizagens alcançadas, deve ser feita, de forma colaborativa, uma síntese das principais ideias trabalhadas ao longo da resolução da tarefa. No final da aula, os alunos refletem individualmente sobre o seu trabalho e o que aprenderam.

Esta tarefa pode ser adaptada pelo docente às características da escola, do meio sociocultural envolvente e do grupo turma.

Dificuldades esperadas e ações do professor:

Os alunos podem revelar algumas dificuldades, nomeadamente em:

- Não terem presente as noções estatísticas estudadas no ensino básico, que permitam calcular a média, a mediana ou construir um diagrama de caule-e-folhas.
- Organizar a informação da Parte II da tarefa, para apresentar os resultados do seu trabalho à turma, ou na escolha dos recursos a utilizar, decidindo se recorrem ou não ao uso da tecnologia.

O professor, sempre que necessário, deve rever os conceitos estatísticos de anos anteriores e ajudar os alunos na utilização da tecnologia (folha de cálculo e/ou calculadora gráfica), dando sugestões ou colocando questões que contribuam para o desenvolvimento do trabalho.

Avaliação:

A avaliação poderá ter um carácter formativo.

Sugere-se que, na Parte II, cada grupo apresente a sua resolução à turma, tendo em conta os aspetos sociais e económicos da desigualdade de rendimentos em Portugal e a valorização da estatística para a resolução da tarefa, justificando as suas opções. Esta apresentação poderá ser o suporte de uma avaliação formativa.

Possíveis Aprofundamentos:

Esta tarefa tem potencialidades para ser utilizada em contexto interdisciplinar ou como contributo para a Cidadania e Desenvolvimento, em Economia, em Geografia, em Área de Integração, entre outras. Em contexto interdisciplinar, pode ser dinamizada uma discussão mais alargada sobre outros tipos de desigualdade, tal

como a desigualdade de género, a qual poderá ser promovida, por exemplo, a partir da visualização do vídeo “Desigualdade de género no olhar das crianças”¹².

¹² Disponível em: <https://www.youtube.com/watch?v=Vblc4GDplkQ>

PROPOSTA DE RESOLUÇÃO

PARTE I

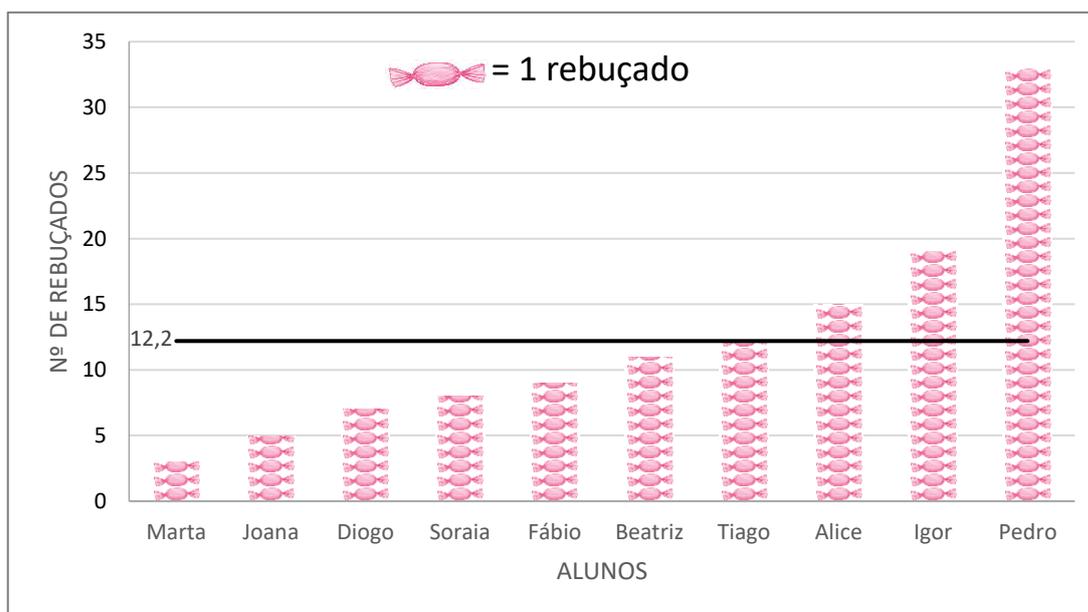
1. A pergunta é de resposta aberta, pelo que poderão surgir várias ideias diferentes. Por exemplo, por observação da tabela, os alunos irão eventualmente notar uma grande diferença entre as quantidades de rebuçados obtidas pelas diversas crianças. Poderão imaginar diversas explicações: alguns dos meninos tiveram sorte e bateram à porta de pessoas generosas que lhes deram pacotes de rebuçados... os menos afortunados bateram às portas, mas não estava ninguém em casa... os mais tímidos não insistiram o suficiente para conseguirem ganhar rebuçados...

Também poderá aparecer a ideia de que o total de rebuçados alcançado pelo grupo (122) foi elevado, apesar da desigualdade entre as coletas individuais. Eventualmente, alguém poderá considerar “estranha” ou até “incompreensível” uma tão grande variação na obtenção de rebuçados pelas várias crianças. Esta é uma ideia que fará sentido salientar, pois aponta para a **desigualdade** ou **grande variabilidade** dos dados daquele conjunto, sobretudo para a existência de um valor substancialmente maior do que os restantes.

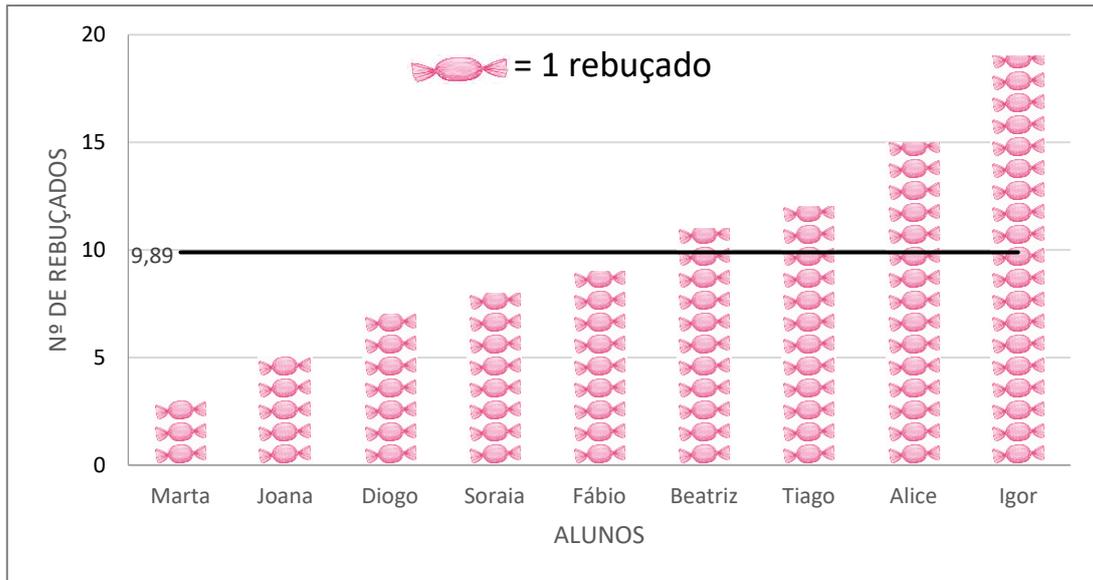
2. A média é facilmente obtida, dado o total de rebuçados, que é igual a 122, e o número de crianças que é 10. A média será:

$$\bar{x} = \frac{122}{10} = 12,2$$

Neste caso, a média não dá uma informação muito útil sobre a distribuição de rebuçados, uma vez que existe um valor atípico, a que damos o nome de *outlier*, que influencia a média. Em média, cada aluno recebeu cerca de 12 rebuçados, mas é evidente que os alunos não foram igualmente premiados com 12 rebuçados, isto é, vários deles ficaram muito abaixo da média e vários ficaram bastante acima da média. A representação dos dados num pictograma (construído no Excel), onde está exibida a linha da média, permite visualizar as distâncias dos dados em relação à média.



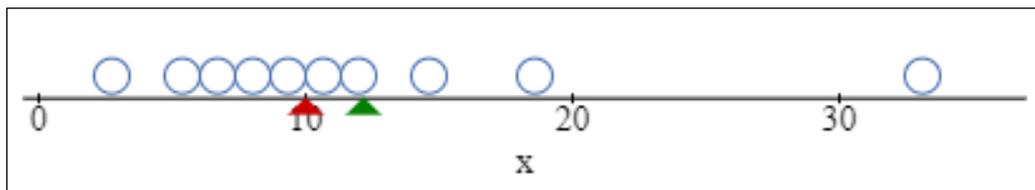
Repare-se que se retirarmos o *outlier* ao conjunto dos dados o valor da média desce para $\bar{x} = 9,89$ e 5 dos dados estão abaixo da média e 4 acima da média, como mostra o pictograma seguinte:



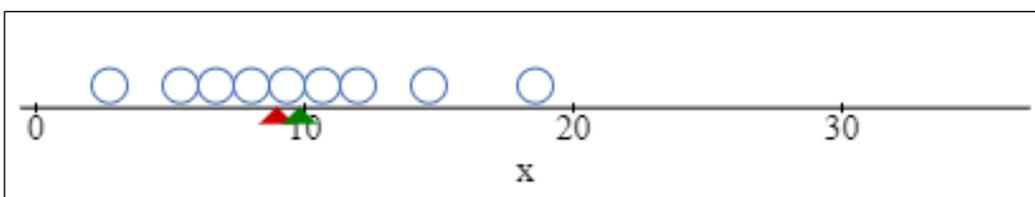
A visualização do efeito do *outlier* sobre a média (média assinalada com seta verde; mediana assinalada com seta vermelha) pode ser apoiada pelos dois diagramas seguintes: (a) com os dados todos e (b) sem o *outlier*.

Nota: Os diagramas foram construídos com o aplicativo *StatCrunch – Mean versus Median*, disponível em <https://www.statcrunch.com/applets/type1&meanmedian>.

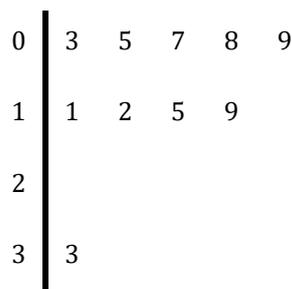
(a)



(b)



3. O diagrama de caule-e-folhas é o seguinte:



Legenda: 0|3 significa 3 rebuçados

A partir do diagrama, verifica-se que a maior concentração de dados surge no intervalo de 0 a 9. Nota-se uma ausência de dados entre 20 e 29. Ocorre um valor atípico (*outlier*) no intervalo de 30 a 39, nomeadamente o valor 33, que provoca uma forte assimetria e, neste caso, uma assimetria à direita.

4. Sendo uma distribuição com um pequeno número de dados, o cálculo da mediana pode ser feito facilmente, com papel e lápis. Os dados já estão ordenados na tabela dada.

3 5 7 8 9 11 12 15 19 33



A mediana separa os dados em duas partes com o mesmo número de elementos. Como o número de dados é par, considera-se que a mediana é a semi-soma dos 2 valores centrais:

$$M_e = \frac{9 + 11}{2} = 10$$

A mediana indica que há metade dos dados com valores menores ou iguais a 10 e a outra metade com valores maiores ou iguais a 10.

No conjunto dos dados em estudo, o valor da média está à direita do valor da mediana, ou seja, $M_e < \bar{x}$. Isto está de acordo com o que se viu anteriormente, isto é, um enviesamento da distribuição para a direita. A média é uma medida de localização do centro da distribuição dos dados sensível a valores atípicos (*outliers*), razão pela qual, neste caso, a média está deslocada para a direita, relativamente à mediana, a qual é apenas condicionada pelos valores centrais da distribuição.

5. Os quartis dividem o conjunto dos dados em 4 partes com aproximadamente 25% dos dados cada uma. O primeiro quartil, Q_1 , e o terceiro quartil, Q_3 , são as medianas de cada uma das partes em que fica dividido o conjunto dos dados pela mediana. Assim, uma vez que a mediana já foi calculada anteriormente, para obter o 1º Quartil e o 3º Quartil, basta calcular a mediana do conjunto 3, 5, 7, 8, 9 e do conjunto 11, 12, 15, 19, 33, respetivamente.

Cálculo do 1º quartil:

Como temos um número ímpar de dados, a mediana é o valor do meio, ou seja, $Q_1 = 7$. Assim, concluímos que há aproximadamente 25% dos dados menores ou iguais a 7 e os restantes dados maiores ou iguais a 7.

Cálculo do 3º quartil:

Da mesma forma que para o 1º quartil, também temos um número ímpar de dados, pelo que a mediana é o valor do meio, ou seja, $Q_3 = 15$. Assim, concluímos que há aproximadamente 75% dos dados menores ou iguais a 15 e os restantes dados maiores ou iguais a 15.

Nota: Se o conjunto de dados tiver um número ímpar de dados, então a mediana é um elemento desse conjunto. Neste caso, colocam-se duas opções para o cálculo dos quartis:

- Ou se considera a mediana como pertencente a ambas as partes em que fica dividido o conjunto dos dados, ou
- Não se considera a mediana pertencente a nenhuma das partes.

Pode-se utilizar qualquer das metodologias.

Concretizando, na situação do número de rebuçados, percebemos que cerca de $\frac{1}{4}$ dos alunos recebeu menos do que ou igual a 7 rebuçados e cerca de $\frac{3}{4}$ dos alunos receberam mais do que ou igual a 7 rebuçados. Além disso, sabemos que cerca de $\frac{3}{4}$ dos alunos receberam menos do que ou igual a 15 rebuçados e cerca de $\frac{1}{4}$ dos alunos recebeu mais do que ou igual a 15 rebuçados.

6. São calculadas as percentagens (com aproximação às centésimas) e é preenchida a nova coluna:

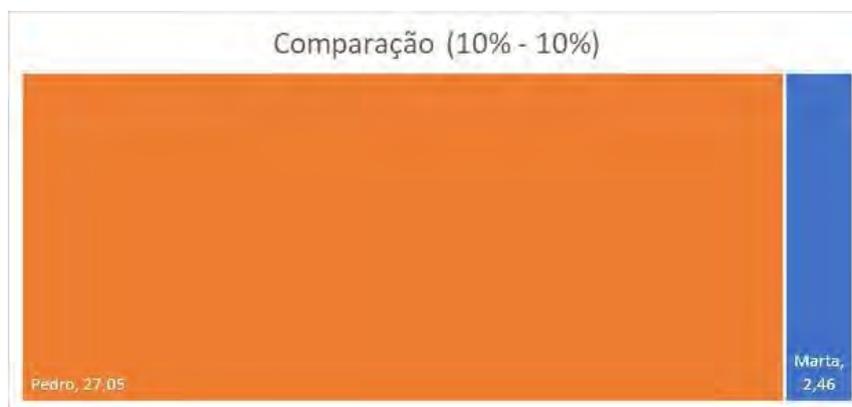
Alunos	Nº de rebuçados	Parcela do total (%)
Marta	3	2,46
Joana	5	4,10
Diogo	7	5,74
Soraia	8	6,56
Fábio	9	7,38
Beatriz	11	9,02
Tiago	12	9,84
Alice	15	12,30
Igor	19	15,57
Pedro	33	27,05

Nota: O total da coluna das percentagens é 100,02% (e não 100%) devido aos arredondamentos efetuados.

7.1. A afirmação é verdadeira. Os 30% mais favorecidos do grupo foram o Pedro, o Igor e a Alice. Eles correspondem aos 3 estratos de 10% que ganharam mais rebuçados. Em conjunto, estes 30% de alunos ganharam $12,30\% + 15,57\% + 27,05\% = 54,92\%$ do total de rebuçados. Este resultado é indicador de uma forte desigualdade na distribuição (em conjunto, 3 alunos ganharam mais de metade do total dos rebuçados).

7.2. Os 10% de alunos mais favorecidos equivalem, neste caso, ao Pedro, que ganhou 27,05% (ou 33 rebuçados). Os 10% de alunos menos favorecidos equivalem à Marta que ganhou apenas 2,46% (ou 3 rebuçados). Fazendo a razão entre os valores do Pedro e da Marta, obtém-se o quociente igual a 11. Portanto, os 10% mais favorecidos ganharam 11 vezes mais rebuçados do que os 10% menos favorecidos.

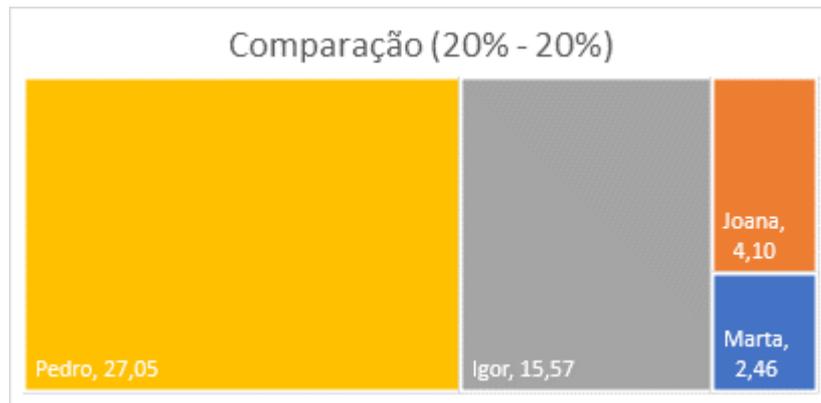
Uma representação útil para a visualização desta situação é um gráfico *Treemap*. Trata-se de um gráfico de hierarquia (disponível no Excel) que mostra os dados representados pelas áreas de retângulos justapostos, permitindo uma rápida comparação entre os valores representados. Tem-se facilmente um destaque visual da proporção entre o maior e o menor valor (no retângulo laranja cabem 11 retângulos azuis).



7.3. A comparação entre os 20% mais favorecidos (Igor e Pedro) e os 20% menos favorecidos (Marta e Joana) é feita de modo análogo. A razão que se obtém é:

$$\frac{15,57 + 27,05}{2,46 + 4,10} = 6,5$$

Portanto, os 20% mais favorecidos do grupo, em conjunto, ganharam 6,5 vezes mais rebuçados do que os 20% menos favorecidos (considerando o gráfico *Treemap*, no total dos retângulos amarelo e cinzento cabe 6,5 vezes o total dos retângulos laranja e azul).



PARTE II

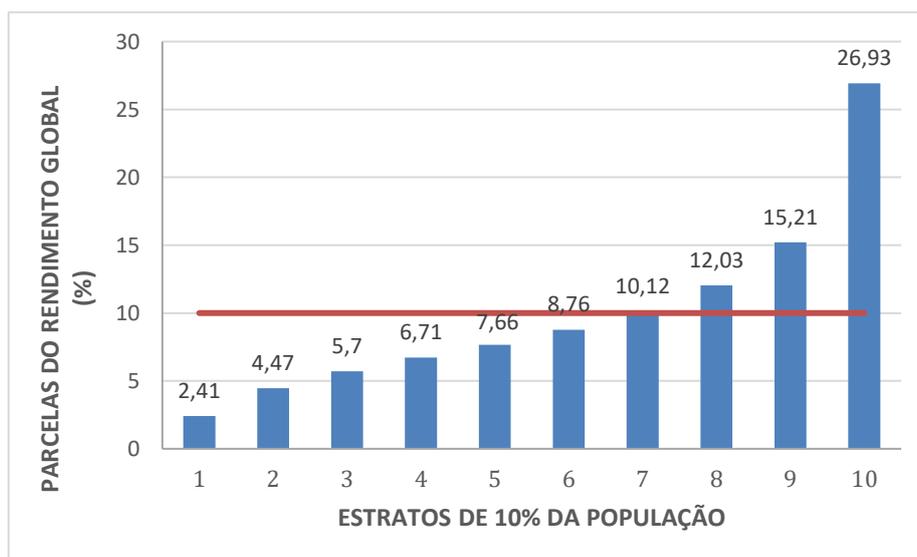
Uma comparação rápida entre a tabela que foi construída com as percentagens de rebuçados obtidos pelas crianças e a tabela que mostra as parcelas de rendimentos dos 10 estratos da população portuguesa, em 2020, permite perceber a semelhança.

Alunos	Nº de rebuçados	Parcela do total de rebuçados (%)
Marta	3	2,46
Joana	5	4,10
Diogo	7	5,74
Soraia	8	6,56
Fábio	9	7,38
Beatriz	11	9,02
Tiago	12	9,84
Alice	15	12,30
Igor	19	15,57
Pedro	33	27,05

Estratos da população (formados por decis)	Parcela do rendimento total (%)
1	2,41
2	4,47
3	5,70
4	6,71
5	7,66
6	8,76
7	10,12
8	12,03
9	15,21
10	26,93

Da análise já feita para a situação da distribuição de rebuçados, podemos deduzir uma forte desigualdade na distribuição de rendimentos da população portuguesa, em 2020. A distribuição de rendimentos é assimétrica para a direita, observando-se a existência do *outlier* com o valor 26,93 que é a percentagem de rendimento do estrato da população com rendimentos mais elevados.

Se houvesse uma distribuição de rendimentos perfeitamente equitativa, então todos os estratos da população ganhariam o mesmo, isto é, cada estrato ganharia 10% do rendimento total. Representando as parcelas ganhas pelos diversos estratos num gráfico com barras e exibindo a reta horizontal correspondente a 10% do rendimento global, podemos visualizar como as parcelas se desviam da média, isto é, como há vários estratos abaixo da média e alguns acima da média. O gráfico obtido é bastante parecido com o que foi construído para a distribuição do número de rebuçados.



A distribuição de rebuçados ganhos pelos 10 alunos no Halloween é comparável ao que se verificou, em 2020, com a distribuição de rendimentos no nosso país, quando se considera a população dividida em 10 estratos.

Tal como na distribuição do número de rebuçados, esta distribuição de rendimentos evidencia uma grande desigualdade e um enviesamento à direita. Observamos que cada um dos 3 estratos correspondentes aos 30% mais ricos ganha mais do que a média nacional de rendimentos. Pelo contrário, cada um dos outros 6 estratos ganha menos do que a média nacional de rendimentos.

Dois indicadores importantes de desigualdade numa distribuição de rendimentos são os quocientes já calculados no caso anterior (também se designam por rácios S90/S10 e S80/S20).

Assim, em 2020, os 10% dos portugueses mais ricos ganhavam aproximadamente 11 vezes mais do que os 10% dos portugueses mais desfavorecidos:

$$\frac{26,93}{2,41} = 11,17$$

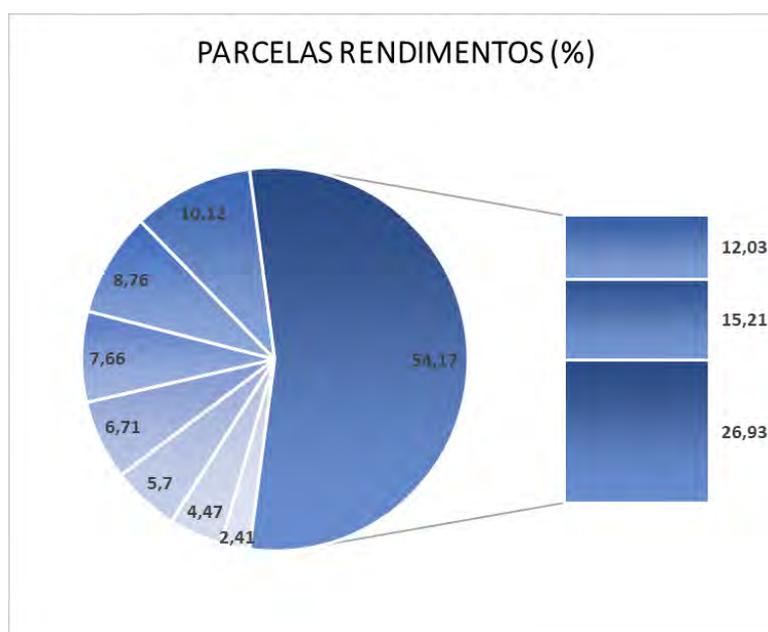
Em termos figurativos, em 2020, os 10% mais ricos da população portuguesa equivalem ao Pedro (com 33 rebuçados) e os 10% mais pobres equivalem à Marta (com 3 rebuçados).

Além disso, os 20% mais ricos da população ganhavam aproximadamente 6 vezes mais do que os 20% menos ricos:

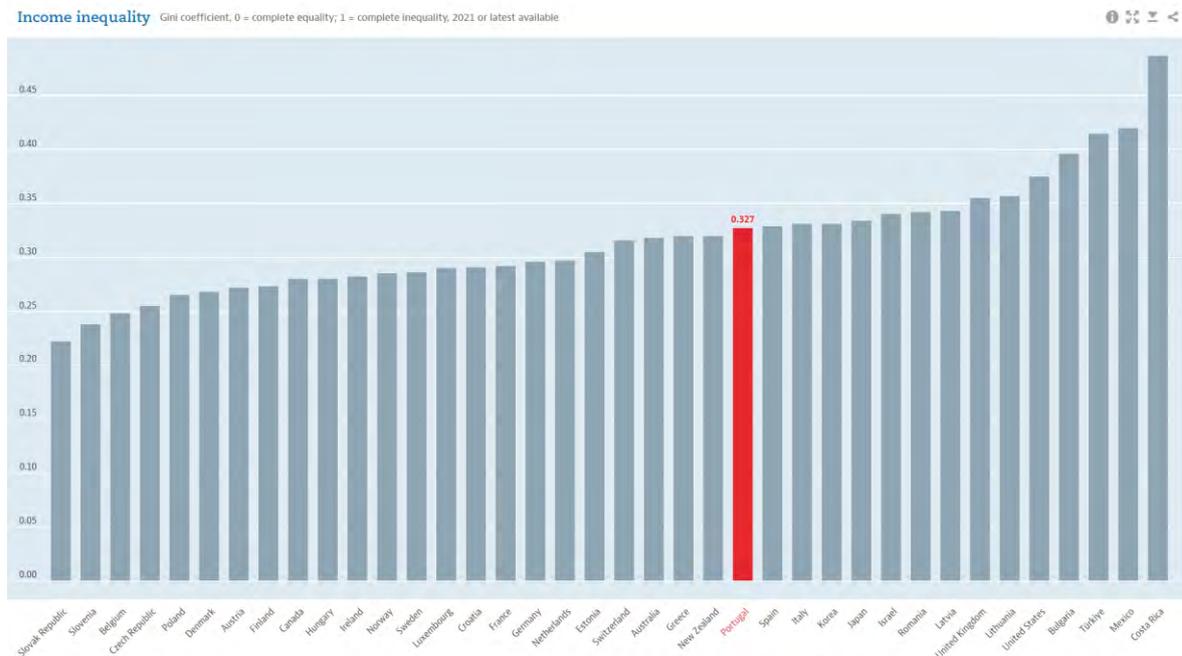
$$\frac{26,93 + 15,21}{2,41 + 4,47} = 6,13$$

Note-se ainda que, em 2020, os 30% mais ricos ganhavam, em conjunto, 54,17% do total do rendimento nacional (equivalem aos 3 alunos que ganharam mais rebuçados: Pedro, Igor e Alice). Isto significa que 30% da população ganhava, conjuntamente, mais do que os restantes 70% juntos. Por outras palavras, em 2020, cerca de 1/3 da população portuguesa, ganhava, conjuntamente, mais do que os outros 2/3 da população, em conjunto.

Um gráfico circular (do tipo barra de circular, obtido no Excel) permite visualizar a desigualdade da distribuição de rendimentos e destacar a proporção entre o conjunto dos três estratos superiores (agrupados num único setor e representados na barra empilhada) e os restantes estratos.



Nota: Será interessante referir a posição de Portugal no *ranking* da desigualdade de distribuição de rendimentos. A OCDE disponibiliza dados que fazem essa comparação, tendo por base o chamado *Índice de Gini*, cujo valor varia entre 0 e 1. Quanto maior é esse índice, maior é a desigualdade da distribuição de rendimentos. Consta-se que, em 2020, a desigualdade de rendimentos em Portugal foi maior do que em muitos outros países, tendo sido ligeiramente menor do que em Espanha, que surge logo a seguir no *ranking*.



Dados da OCDE: Índice de desigualdade (índice de Gini). A vermelho aparece Portugal, com o valor 0.327, em 2020. Obtido de: <https://data.oecd.org/inequality/income-inequality.htm>. Acedido em 25/10/2023.

CONSUMO DE ÁGUA

TAREFA¹³



Imagem reproduzida do Portal da Água / Todos os direitos reservados

“A incerteza e imprevisibilidade da seca e dos seus impactos justificam que se dedique uma atenção permanente a este fenómeno e não apenas uma actuação reactiva a situações extremas”, lê-se no texto da resolução n.º 80/2017 publicada em Diário da República.

Joaquim Poças Martins, por sua vez, recorda que, como estamos diante de um fenómeno global de mudança do clima, temos de nos “preparar para um novo normal”. E o que é essa nova normalidade? “São períodos de seca que se arrastam, um ano seco depois de outro. Praticamente estamos desde 2017 a enfrentar a seca”, diz o especialista em gestão hídrica, para quem uma comissão dedicada à seca acabaria por ser “um trabalho a tempo inteiro”. “Quase até poderíamos criar o Ministério da Seca”, refere o docente da FEUP.

Jornal o Público | 07.05.2023

<https://www.publico.pt/2023/05/07/azul/noticia/portugal-nao-estrategia-agua-usaa-vontadinha-2048760>

Recorreu-se aos dados sobre o volume de água abastecida pela rede pública, recolhidos pela Entidade Reguladora dos Serviços de Águas e Resíduos (ERSAR), disponíveis [AQUI](#)¹⁴. Os dados têm proveniência de fontes como o INE, a ERSAR, a ERSAR Açores e DREM, a partir dos dados reportados pelos Sistemas Públicos Urbanos de Serviços de Águas /Vertente Física e de Funcionamento. Os valores apresentados referem-se ao consumo de água abastecida pela rede pública e correspondem ao volume de água faturada, *per capita*, em cada ano, entre 2015 e 2021.

¹³ Esta tarefa foi adaptada da TarefALEA Consumo de água:

https://www.alea.pt/index.php?option=com_content&view=article&id=955&Itemid=2151&lang=pt

¹⁴ Os dados podem ser descarregados em diversos formatos, incluindo no formato Excel. Note-se que no quadro disponibilizado pelo INE os anos estão ordenados do mais recente para o mais antigo.

PARTE I

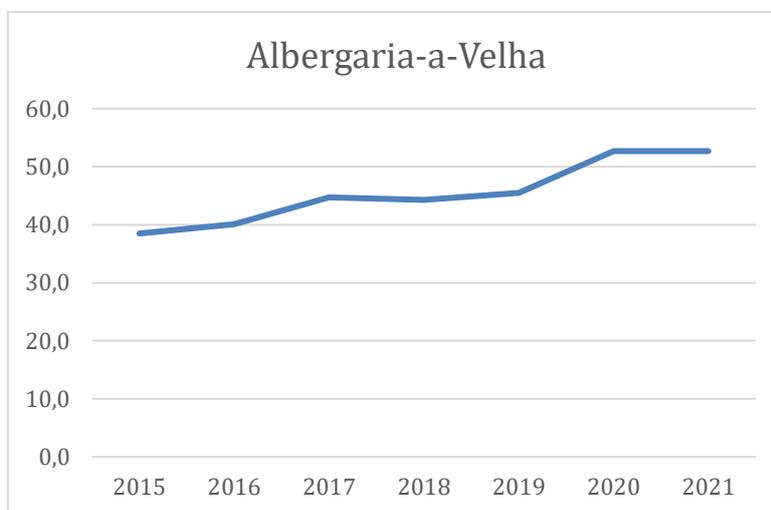
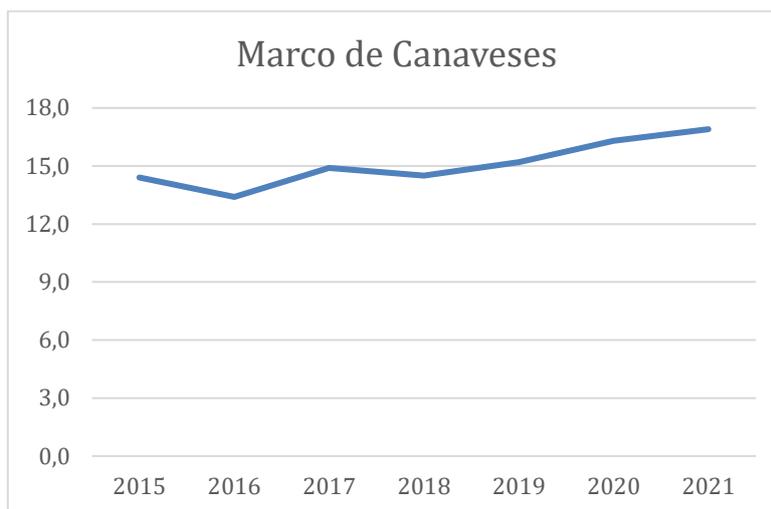
Observa a tabela seguinte na qual está indicado o consumo de água, anual, *per capita* (em m³/hab) de um município de cada distrito de Portugal Continental.

Distrito	Município	2015	2016	2017	2018	2019	2020	2021
		m ³ /ha b						
Viana do Castelo	Vila Nova de Cerveira	61,1	62,6	54,7	55,4	52,6	52,5	50,9
Braga	Fafe	30,2	30,4	31,9	31,5	30,4	31,5	31,7
Vila Real	Ribeira de Pena	49,1	39,1	43,9	42,6	44,4	64,7	43,9
Viseu	Cinfães	14,2	14,8	15	14,5	15,1	25,8	25,7
Porto	Marco de Canaveses	14,4	13,4	14,9	14,5	15,2	16,3	16,9
Bragança	Mogadouro	103,2	100,4	115,3	89,4	100,7	98,4	97,5
Aveiro	Albergaria-a-Velha	38,5	40,1	44,7	44,3	45,5	52,7	52,7
Coimbra	Coimbra	73,7	74,5	77	73,5	75,6	74,9	72,5
Leiria	Ansião	57,8	59,2	59,7	56,4	62,7	56,7	54,5
Castelo Branco	Penamacor	75,5	65,1	67,5	60,7	64	65	64,2
Guarda	Celorico da Beira	53,3	62,6	65,4	64,9	68,3	65,1	68
Lisboa	Lisboa	185,6	174,3	175,5	165,6	158,7	150,2	149,9
Setúbal	Alcácer do Sal	94,3	77,2	71	82	84,2	95,4	99,2
Beja	Alvito	58,8	56,7	61,6	55,8	57,4	68,4	64,2
Santarém	Santarém	67,7	71,6	73,5	71,3	75,5	78	77,3
Portalegre	Monforte	47,6	44,4	44,8	41,6	41,3	39,8	32,4
Évora	Montemor-o-Novo	49,8	48,1	50,3	46,4	49,5	54	52,4
Faro	Vila Real de Sto. António	129,8	105,6	115,2	118,6	107,8	100,8	105

1. Indica a unidade de observação, a variável estatística e a unidade de medida.
2. Qual é o tipo de variável estatística representada na tabela?
3. Relativamente ao ano de 2021, indica os três municípios com maior consumo de água e os três municípios com menor consumo de água.
4. Para cada município, analisa a variação do consumo entre 2015 e 2021 e a percentagem de variação ocorrida relativamente ao ano de 2015. Em qual dos municípios se poderia supor que houve um comportamento mais consciente de consumo da água? E em qual dos municípios se poderia supor o contrário? Explica a tua resposta.
5. Observando a tabela, verifica-se que o consumo de água, em Lisboa, no ano de 2021, foi 149,9 m³/hab. Poderá afirmar-se que no município de Lisboa, em 2021, cada pessoa que vive em Lisboa consumiu, em média, essa quantidade de água? Justifica.

6. De acordo com dados disponíveis no PORDATA¹⁵ sobre a população residente por municípios, em 2021, existiam 49 637 habitantes em Marco de Canaveses e 18 946 em Vila Real de Sto. António. Com base nesta informação, compara os consumos de água anuais dos dois municípios, face à população residente nesse ano.

7. Observa os gráficos seguintes e compara as tendências de evolução dos consumos *per capita* nos dois municípios. Comenta a afirmação: “A evolução do consumo é semelhante nos dois municípios, ao longo do tempo”.



8. Segundo a ONU, 110 litros de água por dia são suficientes para atender às necessidades básicas de consumo e higiene de uma pessoa. Com base nesta informação, indica quantos m³ *per capita* seriam recomendáveis consumir por dia e por ano.

¹⁵ <https://www.pordata.pt/municipios/populacao+residente+total+e+por+grandes+grupos+etarios-390>

PARTE II

De acordo com o teu interesse, seleciona do conjunto de dados “Consumo Água Município”, disponível [AQUI](#), pelo menos, 15 municípios e organiza-os numa tabela, usando a folha de cálculo.

1. Considera os dados relativos ao ano de 2021 e representa-os num diagrama de caule-e- folhas. (Na construção do diagrama de caule-e-folhas podes desprezar a parte decimal do número pois o aspeto gráfico é o mesmo).

1.1. Com base no diagrama, obtém a amplitude dos consumos *per capita*, em 2021.

1.2. Indica uma vantagem do gráfico que obtiveste para descrever os dados.

Nota: No gráfico de caule-e-folhas os dígitos do caule têm ordem mais elevada e os das folhas são os de ordem imediatamente seguinte. Neste caso, 1|6, por exemplo, lê-se 16 m^3 .

2. Para cada um dos municípios, determina a média anual dos consumos *per capita*, entre 2015 e 2021. Como interpretas esses valores?

3. Escolhe os municípios que registaram a maior e a menor média anual.

3.1. Constrói um gráfico de linhas para cada caso, que mostre a evolução do consumo de água, *per capita*, no período de 2015 a 2021. Interpreta o gráfico obtido.

3.2. Indica uma vantagem do gráfico que obtiveste para descrever os dados.

4. Segundo a ONU, 110 litros de água por dia são suficientes para satisfazer as necessidades básicas de consumo e higiene de uma pessoa. Com base nesta informação e nos dados do consumo de água *per capita* do teu município, que conclusões podes tirar e que recomendações podes apresentar?

5. Que aspetos consideras que devem ser estudados para efetuares recomendações sobre o consumo de água nos municípios que selecionaste?

NOTAS PARA O PROFESSOR

A exploração da tarefa **Consumo de Água** é adequada às Aprendizagens Essenciais de Matemática A, Matemática B, MACS e Matemática dos Cursos Profissionais.

A tarefa tem um grande foco no desenvolvimento da Educação para a Cidadania e numa perspetiva de Matemática para Todos, promovendo a comunicação matemática e o desenvolvimento do espírito crítico.

Tema: Estatística

Tópicos e subtópicos:

Organização de dados

- Diagrama de caule-e-folhas

Dados univariados

- Medidas de localização
- Medidas de dispersão
- Propriedades das medidas

Dados bivariados - Gráfico de linhas

Objetivos de aprendizagem:

- Organizar e representar a informação contida em dados quantitativos contínuos em tabelas e interpretá-las.
- Selecionar representações gráficas adequadas para cada tipo de dados identificando vantagens/inconvenientes, relembrando a construção de diagramas de caule-e-folhas.
- Interpretar as medidas de localização, em particular a média (\bar{x}), na caracterização da distribuição dos dados, relacionando-a com as representações gráficas obtidas.
- Interpretar as medidas de dispersão, em particular a amplitude (evidenciada no diagrama de caule-e-folhas).
- Entender que um gráfico de linhas é um caso particular de um diagrama de dispersão, em que se pretende estudar, num par de variáveis, a evolução de uma das variáveis relativamente a outra variável, de um modo geral o tempo, e em que se unem, por linhas, os pontos representados.

Duração prevista: Quatro aulas de 50 minutos.

Materiais e recursos:

- Computador, calculadora gráfica, telemóvel.
- Internet, folha de cálculo.

Estratégias de implementação:

A notícia “Portugal não tem estratégia para a água – mas usa-a à vontade”, publicada no Jornal o Público refere-se a um tema da atualidade que deve fazer parte das preocupações de todos os cidadãos.

Recomendamos que, numa primeira fase da aula, o professor dinamize uma discussão coletiva sobre a importância da água e do seu consumo consciente. Numa segunda fase, os alunos devem ser organizados em grupos para a resolução da tarefa. No trabalho de grupo, todos os membros do grupo devem colaborar para as resoluções apresentadas. Neste sentido, os alunos devem avaliar e comentar as possíveis respostas sugeridas pelos colegas, identificando os pontos fortes e fracos e escolhendo a melhor proposta.

A tarefa permite a revisão de alguns conceitos estatísticos estudados no ensino básico. A resolução da tarefa deverá permitir que os alunos se apropriem dos conceitos estatísticos subjacentes aos tópicos a estudar. A segunda parte da tarefa permite que os alunos selecionem um conjunto de municípios do seu interesse e se debruçam sobre questões do consumo de água nesses locais.

O facto de a segunda parte da tarefa ser mais aberta, pode levar a novas questões relativas ao consumo de água e a diversos conceitos estatísticos.

Na última fase deverão ser apresentados os trabalhos de diferentes grupos e ser promovida a discussão crítica dos resultados em grande grupo. Na discussão global, os alunos irão comparar as diferentes abordagens que viram e usaram. Finalmente, a partir das apresentações realizadas e das aprendizagens deve ser feita, de forma colaborativa, uma síntese das principais ideias trabalhadas ao longo da resolução da tarefa. No final da aula, os alunos refletem individualmente sobre o seu trabalho e o que aprenderam.

Esta tarefa pode ser adaptada pelo docente às características da escola, do meio sociocultural envolvente e do grupo turma.

Dificuldades esperadas e ações do professor:

Os alunos podem revelar algumas dificuldades, nomeadamente em:

- Não terem presente as noções estatísticas estudadas no ensino básico, que permitam calcular a média, a amplitude, representar um diagrama de caule-e-folhas ou um gráfico de linhas.
- Organizar a informação, na Parte II da tarefa.
- Trabalhar com a folha de cálculo/calculadora gráfica, em particular na introdução dos dados, na obtenção dos gráficos de linhas e no cálculo das médias ou amplitudes. A familiarização com a folha de cálculo, usando

ferramentas como a ordenação dos dados, pode ser uma mais-valia na resolução desta tarefa.

O professor, sempre que necessário, deve rever os conceitos estatísticos de anos anteriores e ajudar os alunos na utilização da tecnologia (folha de cálculo e/ou calculadora gráfica), dando sugestões ou colocando questões que contribuam para o desenvolvimento do trabalho.

Avaliação:

A avaliação poderá ter um carácter formativo.

- Sugere-se que, na Parte II, cada grupo entregue a sua resolução por escrito (em princípio cada grupo usará dados diferentes porque irão escolher municípios de acordo com os seus interesses). O professor poderá fornecer um feedback escrito a cada um dos grupos, podendo ainda incluir uma ou duas perguntas sobre o trabalho de cada grupo, ou entregar a cada grupo uma lista de perguntas e orientações para exploração subsequente.
- No final da Parte II, os alunos poderão realizar uma autoavaliação, procedendo a uma reflexão individual sobre as aprendizagens realizadas e estratégias de melhoria.

Possíveis Aprofundamentos:

O desenvolvimento desta temática (consumo de água) pode ser aproveitado para dinamizar um trabalho de projeto, em colaboração com outras disciplinas e em parceria com empresas e instituições locais, como câmaras municipais ou outras. Também podem ser incluídos no estudo deste tema dados de outras regiões do mundo (exemplo de um local com dados sobre a Europa: <https://www.eureau.org/resources/publications/eureau-publications>).

relativamente ao consumo inicial, fazendo a divisão da diferença calculada pelo valor do consumo em 2015. Nas fórmulas introduzidas, foram usadas as referências das células da tabela inicial. Observa-se, então, que o município que teve o maior decréscimo relativo foi o de *Monforte*, onde houve uma diminuição de 32% no consumo *per capita*. Isto poderá sugerir a hipótese de que esteja a haver maior consciencialização no consumo. Em sentido contrário, está o município de *Cinfães*, que apesar de ser um dos municípios com consumo mais baixo em 2021 (cf. Questão 3), registou um aumento percentual de 81%, entre 2015 e 2021. É de notar que poderão existir justificações para este facto que não signifiquem apenas desperdício ou falta de consciência no uso da água.

Município	Diferença entre valor de 2021 e valor de 2015	Porcentagem	Variação de 2015 para 2021
Vila Nova de Cerveira	-10,2	-17%	DECRESC
Fafe	1,5	5%	CRESC
Ribeira de Pena	-5,2	-11%	DECRESC
Cinfães	11,5	81%	CRESC
Marco de Canaveses	2,5	17%	CRESC
Mogadouro	-5,7	-6%	DECRESC
Albergaria-a-Velha	14,2	37%	CRESC
Coimbra	-1,2	-2%	DECRESC
Ansião	-3,3	-6%	DECRESC
Penamacor	-11,3	-15%	DECRESC
Celorico da Beira	14,7	28%	CRESC
Lisboa	-35,7	-19%	DECRESC
Alcácer do Sal	4,9	5%	CRESC
Alvito	5,4	9%	CRESC
Santarém	9,6	14%	CRESC
Monforte	-15,2	-32%	DECRESC
Montemor-o-Novo	2,6	5%	CRESC
Vila Real de Sto. António	-24,8	-19%	DECRESC

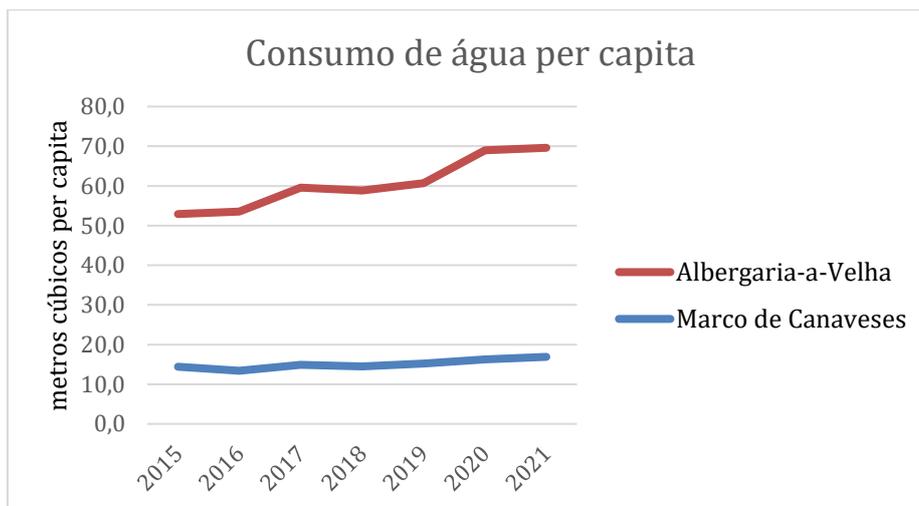
5. Não. É muito importante ter em atenção o significado da variável “água consumida *per capita*”. De acordo com o INE, esta variável tem por base a quantidade de água distribuída e/ou faturada no município, que é dividida pelo número de residentes. Ora, a água distribuída inclui o consumo doméstico ou residencial, mas também o consumo não-residencial, isto é, a água consumida no comércio, indústria e instituições (como escolas, hospitais, etc.). O consumo de água de cada residente em Lisboa foi, portanto, inferior a 149,9 m³. Bastará pensar no volume de água gasto pelos turistas e pelos hotéis (entre outros), que não são residentes. Portanto, o volume de água a dividir pelo número de residentes corresponde a um número que está inflacionado.

6. Começamos por organizar a informação que temos, numa tabela. Esta organização não é obrigatória, mas ajuda a sistematizar os dados para a resolução do problema.

Dados 2021	Consumo de água (<i>per capita</i>)	População residente	Consumo de água total (município)
Marco de Canavezes	16,9	49 637	838 865 (= 16,9 × 49637)
Vila Real de Sto. António	105	18 946	1 989 330 (= 105 × 18946)

Repare-se que, embora a população de Vila Real de Sto. António seja muito inferior à população de Marco de Canavezes, o consumo global anual de água é muito superior. De facto, a razão entre as populações indica-nos que a população de Marco de Canavezes é cerca de 2,6 vezes superior à população de Vila Real de Sto. António. Porém, a razão entre os consumos globais dos dois municípios mostra que o de Vila Real de Sto. António é 2,3 vezes superior ao de Marco de Canaveses. Um possível fator a contribuir para este facto pode ser, como já referido, o turismo nacional e internacional na região do Algarve, especialmente durante os meses de verão, que gera aumento do consumo de água; porém os turistas não são considerados população residente.

7. Numa primeira observação, em ambos os casos, é visível uma tendência de crescimento ao longo do tempo, embora com algumas flutuações. Por exemplo, há uma subida do consumo de 2016 para 2017, em ambos os municípios, seguida de uma pequena descida de 2017 para 2018 e depois o crescimento é retomado nos dois casos. Uma leitura superficial poderia levar a concluir que a evolução nos dois municípios é semelhante. No entanto, é essencial ter em conta que os dois gráficos apresentam escalas diferentes no eixo vertical. Isso tem uma grande influência na leitura do gráfico. Por exemplo, o consumo máximo em Marco de Canaveses é muito menor do que o consumo mínimo de Albergaria-a-Velha. Além disso, também se pode perceber que o acréscimo que se está a observar em Marco de Canaveses é mais lento (cada intervalo na escala é 3 unidades) do que o de Albergaria-a-Velha (cada intervalo na escala é 10 unidades). Uma forma de observar a diferença é representar as duas linhas no mesmo gráfico, sendo a escala igual para as duas (como no gráfico seguinte).

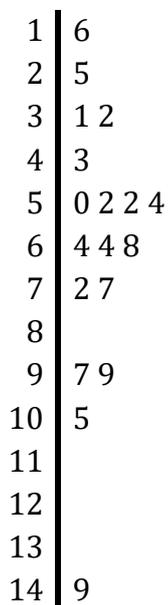


8. Sabemos que 1 litro de água é igual a 1 dm^3 . Então, 110 litros de água correspondem a 110 dm^3 . Uma vez que 110 litros representa o consumo diário, podemos afirmar que o consumo anual seria $110 \times 365 = 40\,150$ litros de água, ou seja, $40\,150 \text{ dm}^3$.

Fazendo a conversão de dm^3 para m^3 , então o consumo indicado pela ONU como suficiente seria de $40,15 \text{ m}^3$, *per capita*, por ano. Repare-se que em 2021, apenas 4 municípios da nossa tabela (ordenada por consumo) apresentam um consumo inferior a esse valor: *Marco de Canaveses, Cinfães, Fafe e Monforte*.

PARTE II

1. Na resolução da **Parte II** da tarefa, os concelhos aqui utilizados foram os da tabela usada na **Parte I** (a título de exemplo). Apresenta-se em seguida o diagrama de caule-e-folhas para os dados de 2021.



Legenda: 1|6 lê-se 16 m³

1.1. A amplitude dos consumos *per capita*, nos municípios representados é dada pela diferença entre o valor máximo (149) e valor mínimo (16). Então, a amplitude será $149 - 16 = 133 \text{ m}^3$ (ou seja, 133 000 litros). Assumindo que a capacidade de uma banheira é cerca de 200 litros, então a amplitude do consumo *per capita* entre municípios corresponde a 665 banheiras cheias de água.

1.2. O diagrama de caule-e-folhas permite ter uma representação rápida, de dados numéricos univariados, em que os dados estão agrupados, mas permitindo reconstituir os dados originais. Ordenando as folhas penduradas em cada caule, a partir desta representação é fácil encontrar o máximo e o mínimo e, com isso, a amplitude dos dados. Apresenta os dados agrupados em classes de igual amplitude e permite ver aspetos da distribuição de dados, tais como, simetria, enviesamento, zonas de maior ou menor concentração ou com lacunas. Reparámos, por exemplo, que a representação anterior apresenta um enviesamento para a direita, pois há maior concentração nas primeiras 7 classes do que nas últimas 7 classes, que apresentam algumas lacunas. Nota-se a existência de um valor, o 149 m³, que sobressai dos restantes, podendo identificar a existência de um *outlier*. A maior concentração dos dados ocorre no intervalo [50, 80[.

2. Usando a folha de cálculo, apresentam-se na tabela seguinte as médias dos consumos por município, entre os anos de 2015 a 2021.

Podemos observar que a média de consumos no conjunto dos anos é bastante diferente de município para município, reforçando a ideia de que há municípios bastante distintos no consumo de água *per capita*. Os municípios com médias mais baixas são novamente (comparar¹⁶ com os dados do ano de

¹⁶ Está aqui a ser feita uma comparação porque se usaram os mesmos municípios da Parte I da tarefa. Porém, os alunos irão provavelmente escolher outros municípios, de acordo com o seu interesse, pelo que poderão não surgir comparações com os resultados obtidos na Parte I.

2021): Marco de Canaveses, Cinfães e Fafe. Os municípios com médias mais altas são: Lisboa, Vila Real de Sto. António e Mogadouro.

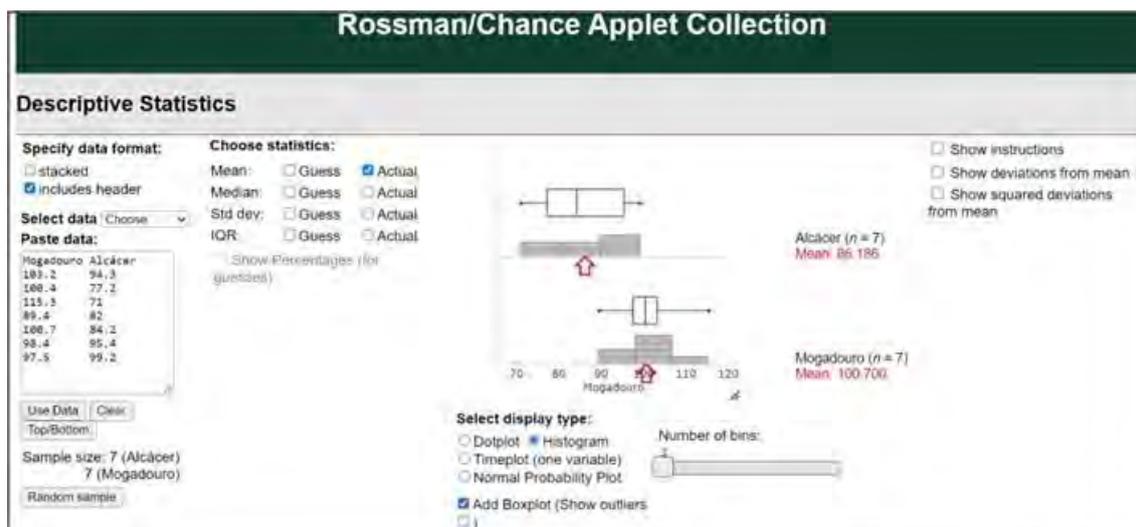
=MÉDIA(E5:K5)

	A	B	C	D	E	F
29						
30						
31						
32				Município	Média entre 2015 e 2021	
33				Vila Nova de Cerveira	55,7	
34				Fafe	31,1	
35				Ribeira de Pena	46,8	
36				Cinfães	17,9	
37				Marco de Canaveses	15,1	
38				Mogadouro	100,7	
39				Albergaria-a-Velha	45,5	
40				Coimbra	74,5	
41				Ansião	58,1	
42				Penamacor	66,0	
43				Celorico da Beira	63,9	
44				Lisboa	165,7	
45				Alcácer do Sal	86,2	
46				Alvito	60,4	
47				Santarém	73,6	
48				Monforte	41,7	
49				Montemor-o-Novo	50,1	
50				Vila Real de Sto. António	111,8	
51						

É interessante notar¹⁷ que o município de Mogadouro não estava no grupo dos 3 municípios com maior consumo em 2021. Em vez de Mogadouro, aparecia Alcácer do Sal. Isto mostra que a média em Mogadouro é afetada por valores mais elevados ocorridos em anos anteriores a 2021.

Recorrendo a um aplicativo disponível na Internet (em Inglês) podemos comparar rapidamente as médias e as distribuições dos consumos (na forma de histogramas e diagramas de extremos e quartis). Essa comparação está ilustrada abaixo, entre os municípios de Mogadouro e Alcácer do Sal. Observe-se a diferença nas médias e nos histogramas, em particular.

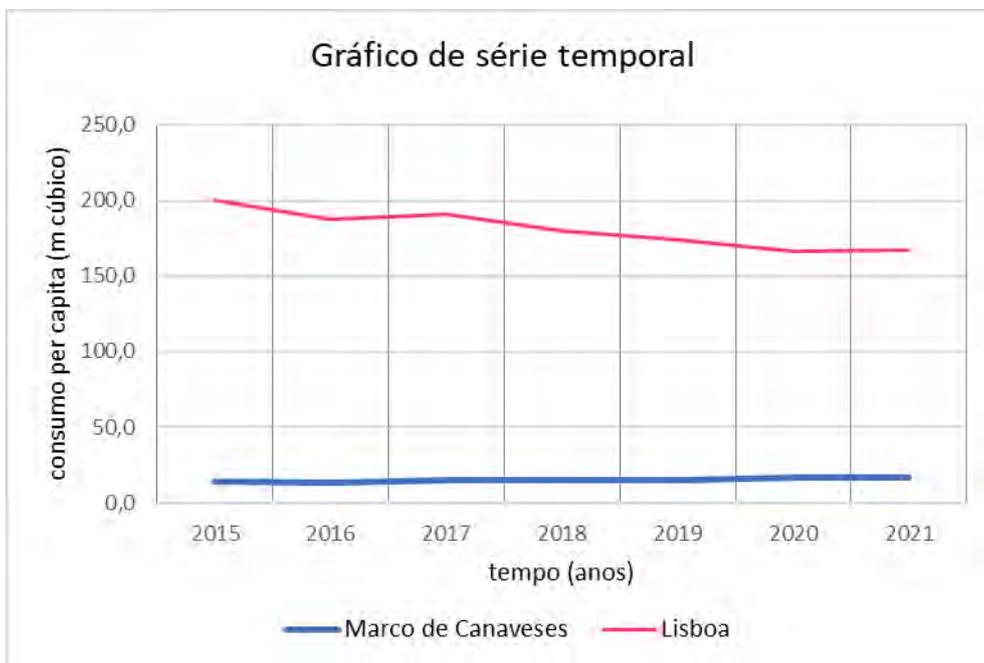
Nota: O aplicativo *Descriptive Statistics* está disponível em <http://www.rossmanchance.com/applets/2021/descstats/Dotplot.htm> Aqui podemos colar (no retângulo da esquerda “Paste Data”), os dados copiados do Excel sob a forma de colunas, devendo os números ter ponto decimal e não vírgula decimal.



¹⁷ Idem nota anterior.

3.

3.1. O gráfico de linhas seguinte mostra a evolução temporal do consumo de água, por habitante, nos municípios que registaram a menor e a maior média de consumos entre 2015 e 2021, que foram, respetivamente, *Marco de Canaveses* e *Lisboa*.



3.2. O gráfico de linhas é uma representação gráfica apropriada para visualizar, em dados bivariados, a evolução de uma das variáveis em relação à outra, que normalmente se considera o tempo. Assim, pode ser designado por gráfico de tempo ou gráfico de série temporal porque exhibe os valores de uma variável em relação ao tempo. Neste caso, são exibidos os consumos *per capita* do município (no eixo vertical) em relação ao tempo, em anos (no eixo horizontal).

O gráfico de linhas permite-nos facilmente perceber se há um crescimento ou decréscimo do consumo *per capita* e a sua ordem de grandeza, facilitando a identificação de padrões ou tendências nos dados ao longo do tempo. Assim, observa-se um padrão de decréscimo do consumo *per capita* em Lisboa, com uma tendência para estabilização nos dois últimos anos. Em comparação, no caso de Marco de Canaveses, a variação ao longo do tempo é quase indistinta, pelo que o consumo *per capita* se mantém relativamente reduzido ao longo dos anos.

4. Supondo que o meu município é Coimbra, podemos observar os consumos médios *per capita*, entre 2015 e 2021, na tabela inicial.

	D	E	F	G	H	I	J	K
4		m ³ /hab						
5	Vila Nova de Cerveira	61,1	62,6	54,7	55,4	52,6	52,5	50,9
6	Fafe	30,2	30,4	31,9	31,5	30,4	31,5	31,7
7	Ribeira de Pena	49,1	39,1	43,9	42,6	44,4	64,7	43,9
8	Cinfães	14,2	14,8	15	14,5	15,1	25,8	25,7
9	Marco de Canaveses	14,4	13,4	14,9	14,5	15,2	16,3	16,9
10	Mogadouro	103,2	100,4	115,3	89,4	100,7	98,4	97,5
11	Albergaria-a-Velha	38,5	40,1	44,7	44,3	45,5	52,7	52,7
12	Coimbra	73,7	74,5	77	73,5	75,6	74,9	72,5

Sabendo que 110 litros por dia correspondem a cerca de $40,15 \text{ m}^3$ por ano (calculado na pergunta 6 da Parte I), então em qualquer dos anos o consumo *per capita* foi quase o dobro da quantidade de água necessária para cada pessoa.

5. Nos municípios em que os valores apresentam uma grande discrepância em relação a $40,15 \text{ m}^3$, importa perceber quais serão as razões subjacentes a consumos tão elevados. Sabemos que tal pode estar relacionado com questões, como o turismo, entre outras.

Se os problemas identificados estiverem relacionados com as infraestruturas dos municípios, será necessário que se tomem medidas para resolver as situações.

É importante fomentar a consciência dos cidadãos para que estes analisem criticamente os consumos de água nas suas residências. Isto pode ser promovido através de campanhas nacionais¹⁸ ou locais, bem como de notícias publicadas nos media, como a que dá mote a esta tarefa, ou pela educação escolar.

¹⁸ Veja-se o exemplo da campanha “Água é vida – não a desperdice” promovida pelo Grupo Águas de Portugal e a Agência Portuguesa do Ambiente. Mais informações em: <https://portaldaagua.pt/materiais-de-comunicacao/agua-e-vida/>

QUANTO ESTICA UM ELÁSTICO?

TAREFA



Colocamos-te o desafio de responder à seguinte questão: Que comprimento atinge um elástico de secretária, como o da imagem acima, ao ser esticado até rebentar?

1. Parece-te que esta questão é uma questão estatística? Justifica a tua resposta.
2. Para esta tarefa, tens à disposição um pacote grande de elásticos de secretária, todos do mesmo tipo. Comenta cada uma das decisões seguintes para responder ao desafio:
 - i) Testar um elástico do pacote;
 - ii) Testar alguns elásticos do pacote;
 - iii) Testar todos os elásticos do pacote.
3. Aqui tens instruções para um trabalho experimental em que se estica o elástico até ao ponto de rotura e se obtém o comprimento atingido.

Material necessário e montagem:

Elásticos de secretária

1 régua de 50 cm

1 mola de orelhas pequena

1 clip forte

1 fita de pescoço com mosquetão (porta-chaves) ou um pedaço de cordel forte.



Observa a imagem seguinte para a realização da montagem.



Fixa a mola de orelhas à extremidade da régua onde começa a graduação. Baixa a orelha da mola de modo a ficar deitada sobre a régua. Na ponta livre dessa orelha enfia um clip. Na extremidade livre do clip enfia o elástico. Prende a outra ponta do elástico ao mosquetão da fita (ou usa um cordel, dando um nó). Com a régua bem firme sobre uma mesa, puxa a fita continuamente. O elástico vai esticando ao longo da régua (entre o clip e o mosquetão) e chega a um ponto em que se parte. Regista o valor do comprimento atingido pelo elástico quando este rebenta (arredonda ao meio centímetro mais próximo).

Sugestão: usa o telemóvel para gravar o teste em vídeo, podendo a visualização do vídeo (em baixa velocidade) ajudar a obter um valor mais preciso do comprimento máximo atingido pelo elástico.

Nota: O uso da fita para puxar o elástico torna mais fácil aplicar a força de tração que estica o elástico. Aconselha-se precaução, pois a quebra do elástico pode provocar um efeito de chicotada.

a) Organiza o processo de obtenção dos dados experimentais.

- Sugere-se que utilizes uma amostra de 30 elásticos.
- Repete, de forma tão idêntica quanto possível, o processo de esticar cada elástico até partir.
- Regista os valores obtidos (tendo em atenção que a medição feita com a régua está longe de ser rigorosa). Os valores que obtiveres devem ser entendidos como aproximações dos valores reais.

b) Faz uma análise estatística dos dados recolhidos, considerando a população, a amostra e a variável em estudo. Sugere-se representar os dados graficamente, calcular medidas de localização e de dispersão, elaborar o diagrama em caixa de bigodes e verificar a existência de *outliers*, entre outros...

c) Elabora um **relatório escrito** que descreva o processo realizado experimentalmente e a análise estatística efetuada, incluindo todos os cálculos, gráficos, imagens, ecrã da calculadora, excertos de ficheiros produzidos no computador, etc. Apresenta as conclusões a que chegaste relativamente ao desafio que te foi proposto.

NOTAS PARA O PROFESSOR

A exploração da tarefa – **Quanto estica um elástico?** – é adequada às Aprendizagens Essenciais de Matemática A, Matemática B, MACS e Matemática dos Cursos Profissionais.

Esta tarefa requer que os alunos se envolvam ativamente em trabalho experimental para obter os dados. Por ser uma tarefa aberta, a recolha, organização e representação dos dados está a cargo dos alunos.

Tema: Estatística

Tópicos e subtópicos:

Problema estatístico

População, amostra e variável

- Fases de um procedimento estatístico

Dados univariados

Objetivos de aprendizagem:

- Conhecer e interpretar situações do mundo que nos rodeia em que a variabilidade está presente.

Duração prevista: Três aulas de 50 minutos.

Materiais e recursos:

- Elásticos de secretária, 1 régua de 50 cm, 1 mola de orelhas pequena, 1 clip forte, 1 fita de pescoço com mosquetão (porta-chaves) ou um pedaço de cordel forte.
- Computador, calculadora gráfica, telemóvel.
- Folha de cálculo.

Sugestões práticas para o trabalho experimental:

Sugere-se que o pacote de elásticos a ser fornecido à turma tenha, pelo menos, entre 100 e 200 elásticos (existem pacotes de 100 g no mercado, bem como pacotes de 1 kg). Um pacote de 1 kg garante um número de elásticos bastante grande.

A escolha do tipo de elásticos é importante (cada tipo é designado por um N°) pois queremos elásticos que não estiquem mais do que os 50 cm da régua. Elásticos do tipo $N^\circ 16$ ou $N^\circ 14$ serão aconselháveis. Quanto maior é o N° , maior é o comprimento do elástico.

Propor que, na turma, o trabalho experimental seja realizado com uma ou duas amostras. No caso de duas amostras, por exemplo, metade da turma poderá estudar uma amostra e a outra metade estudar a outra amostra. Recomenda-se que, para a

mesma amostra, seja o mesmo aluno a esticar os elásticos, de modo que a experiência seja realizada em condições semelhantes para todos os elásticos da amostra (a força aplicada em cada elástico deve ser, tanto quanto possível, a mesma e alunos diferentes terão forças diferentes).

Sugere-se que a dimensão da(s) amostra(s) seja 30 por razões que se prendem com a inferência estatística¹⁹.

Estratégias de implementação:

Recomendamos que o professor organize os alunos em grupos de três de modo a permitir a cada grupo executar, filmar e registar os dados.

Deve ser dada liberdade aos grupos para escolherem a forma de organização e representação dos dados, selecionando a informação estatística pertinente e adequada para a elaboração do relatório. Este relatório será o reflexo das intenções de cada grupo de alunos que elaborou o estudo estatístico. As representações estatísticas obtidas por cada grupo de alunos terão por base um conjunto de dados experimentais diferentes e serão o produto de reflexões distintas sobre diferentes análises. Desta forma, as várias conclusões dos grupos de alunos poderão servir de suporte para uma discussão global, no final da tarefa.

Será expectável que cada grupo obtenha valores diferentes para a média e será interessante, no grupo turma, estudar a distribuição dessas médias.

Na fase final deve ser realizada uma síntese das aprendizagens realizadas e das principais conclusões da tarefa.

Dificuldades esperadas e ações do professor:

Os alunos podem revelar dificuldades, nomeadamente ao nível da organização e escolha das representações estatísticas mais adequadas aos dados obtidos.

O relatório da tarefa apresenta uma dificuldade adicional, na medida em que os alunos irão realizar uma experiência para obter os dados e terão de selecionar as representações estatísticas para procederem à análise de dados.

Esta tarefa, por ser aberta, é bastante desafiante para o professor na medida em que cada grupo de alunos irá obter um conjunto de dados diferente. Assim, cabe ao professor orientar os alunos, mediante a colocação de questões que os leve a refletir sobre os dados, assim como sugerir diferentes representações para o mesmo conjunto de dados de modo a tornar mais evidentes situações como a existência de *outliers*²⁰.

¹⁹ Mais informação acerca da dimensão da amostra pode ser obtida na seguinte ActivALEA:

<https://www.alea.pt/images/topicos/activalea/pdf/ActivAlea29.pdf>

²⁰ A noção de *outlier* e a sua representação através de um diagrama em caixa de bigodes estão descritas na seguinte ActivALEA: https://www.alea.pt/images/topicos/activalea/pdf/ActivAlea20_revista.pdf

Avaliação:

A avaliação pode ter um carácter sumativo com atribuição de uma classificação ao relatório escrito.

Esta tarefa constituiu uma oportunidade para o professor seleccionar duas ou três competências presentes no Perfil dos Alunos à Saída da Escolaridade Obrigatória e para as avaliar ao longo do processo de resolução da tarefa.

Sugestão para uma extensão da tarefa:

Pode ser sugerida aos alunos uma ampliação do estudo, em que se repete o trabalho experimental com outro pacote de elásticos do mesmo tipo, mas de uma marca diferente. Nesse caso, poderá ser feito um estudo comparativo entre as duas marcas.

PROPOSTA DE RESOLUÇÃO

1. Sim, esta é uma questão estatística porque é de esperar que o comprimento atingido pelo elástico até rebentar não seja igual para todos os elásticos do pacote. Estamos a prever a existência de variabilidade porque, embora os elásticos de secretária sejam fabricados com uma dada matéria-prima e um determinado processo industrial²¹, quando chegam ao consumidor não são exatamente iguais. Há diversos fatores ambientais que influenciam o seu comportamento. Assim, vamos considerar o *alongamento máximo* do elástico de secretária como uma variável. O estudo desta variável implicará a recolha de dados.

2. O pacote de elásticos fornecido tem um grande número de elásticos. Se pretendemos estudar o alongamento máximo que é atingido pelos elásticos, pressupondo que existirá variabilidade, então:

i) Testar apenas um elástico do pacote não é uma decisão adequada, pois nada se poderia saber acerca da população de elásticos quanto ao alongamento máximo que é atingido, em média.

ii) Testar alguns elásticos é uma decisão correta porque nos permitirá fazer o estudo de uma população através de uma amostra, o que se faz normalmente, na vida real.

iii) Testar todos os elásticos do pacote não é uma decisão adequada, pois o número de elásticos é muito grande, o que tornaria o processo moroso e pouco prático, acabando por destruir, desnecessariamente, a população.

3. Para a elaboração do relatório, adotámos a seguinte estrutura: I. Introdução; II. Descrição do problema; III. Procedimento experimental; IV. Resultados; V. Análise e interpretação; VI. Conclusões; VII. Referências bibliográficas; VIII. Anexos (caso existam).

Segue-se uma proposta de relatório do estudo realizado.

Relatório

I. Introdução

O estudo estatístico refere-se à capacidade de alongamento de elásticos de secretária, ao serem esticados até rebentarem.

Uma breve pesquisa sobre a questão da elasticidade dos materiais permitiu saber que a elasticidade é uma propriedade mecânica dos materiais que se pode descrever por meio de vários indicadores. Entre esses, está o *Alongamento na Rotura* que quantifica a deformação máxima que um material suporta até quebrar (Caetano, 2022). É definido, em percentagem, pela razão entre a variação do comprimento da amostra e o seu comprimento inicial. Por exemplo, para uma amostra com um comprimento inicial de 100 mm e um comprimento final de 500 mm, o alongamento na rotura é de $400/100 = 400\%$. Com base nesta e noutras características são avaliadas potenciais aplicações do material.

II. Descrição do Problema

O problema proposto consiste em saber qual é o alongamento máximo atingido pelos elásticos de secretária, de certo tipo, quando esticados até ao ponto de rutura.

O estudo estatístico incide sobre a população de elásticos de secretária, do tipo N^o 16, contidos num pacote grande (de 1 kg). Foram selecionados, aleatoriamente, 30 elásticos do pacote, que constituem a amostra. A variável em estudo, *Alongamento máximo atingido pelo elástico*, é de tipo quantitativo contínuo. Os resultados da observação da variável sobre os elásticos selecionados para a amostra constituem os nossos dados, também de tipo quantitativo e contínuo. A unidade de medida é o centímetro.

²¹ Uma breve explicação do processo de fabrico industrial de elásticos de borracha é dada no artigo *Rubber Bands: The Making*. Disponível aqui (em Inglês): <https://www.industrialrubbergoods.com/articles/the-making-of-rubber-bands.html>

III. Procedimento Experimental

O fabricante “Bounce Rubber Bands” apresenta uma tabela de tamanhos de elásticos²², segundo a qual um elástico Nº 16 tem um comprimento de 60 mm e uma largura de 1.5 mm. Assume-se que os elásticos do tipo Nº 16 existentes no pacote cumprem essa especificação.

O material usado para observar o alongamento máximo de cada elástico da amostra foi o seguinte: 1 régua graduada de 50 cm; 1 mola de orelhas pequena; 1 clip metálico; 1 fita de pescoço com mosquetão de metal; 1 luva de cozinha acolchoada (para proteção).

Na montagem feita, de acordo com as instruções, a posição da extremidade do clip que prendia o elástico correspondia a 5,5 cm na régua. O comprimento do elástico (esticado, de forma reta) foi medido a partir desse ponto.

Para segurar a régua, na extremidade em que se fixou a mola, usou-se a pressão manual. Protegeu-se a mão com uma luva de cozinha acolchoada contra a eventual chicotada do elástico ao rebentar (figura 1).



Figura 1. Elástico preso entre o clip e o mosquetão, no início da experiência

O elástico foi enfiado cuidadosamente no clip e no mosquetão. A seguir, foi puxada a fita que segura o mosquetão, ao longo da régua, esticando o elástico até este partir. A experiência com cada um dos elásticos foi gravada em vídeo, usando um telemóvel. O vídeo foi visionado posteriormente em velocidade reduzida ($\frac{1}{4} \times$) e foi percorrida a linha do tempo para identificar o ponto em que o elástico quebrou. Fez-se a leitura do comprimento atingido sobre a régua, em centímetros, com arredondamento à meia unidade mais próxima. Foi registado o valor obtido (veja-se a figura 2, correspondente ao momento do vídeo imediatamente antes da rotura, em que o valor lido foi 47,0 cm). Para obter o comprimento máximo atingido pelo elástico, calculou-se a diferença entre 47,0 cm e 5,5 cm (posição da extremidade fixa do elástico).



Figura 2. Imagem do momento (obtido em vídeo) imediatamente antes da rotura do elástico

Repetiu-se a experiência para cada um dos 30 elásticos da amostra, realizando os mesmos passos e puxando a fita de forma idêntica em cada um dos testes.

²² Tabela disponível em:

https://static1.squarespace.com/static/63e2faddb2631c0a01e3282f/t/651f877402fa952b1f278051/1696565112059/Rubber+Band+Sizing+Chart_2017_FINAL.pdf

IV. Resultados

Foram registados os resultados das 30 experiências com os elásticos da amostra. Dois dos elásticos partiram-se imediatamente sem terem esticado, o que levou a desprezar esses dois resultados. Assim, registaram-se 28 valores de alongamento máximo correspondentes a 28 elásticos. Os dados estão apresentados na tabela seguinte.

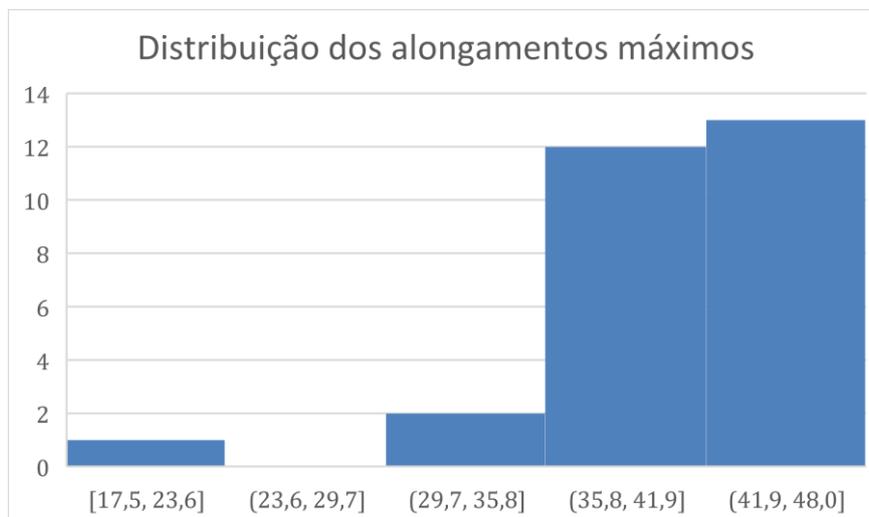
(n = 28)	
Alongamento máximo (cm)	
40,5	43,5
41,0	38,5
43,0	44,0
40,5	41,5
41,5	42,5
34,5	41,0
43,0	45,0
44,0	39,5
42,0	41,5
36,5	43,0
42,0	34,0
44,5	37,0
17,5	42,5
40,5	42,0

V. Análise e Interpretação

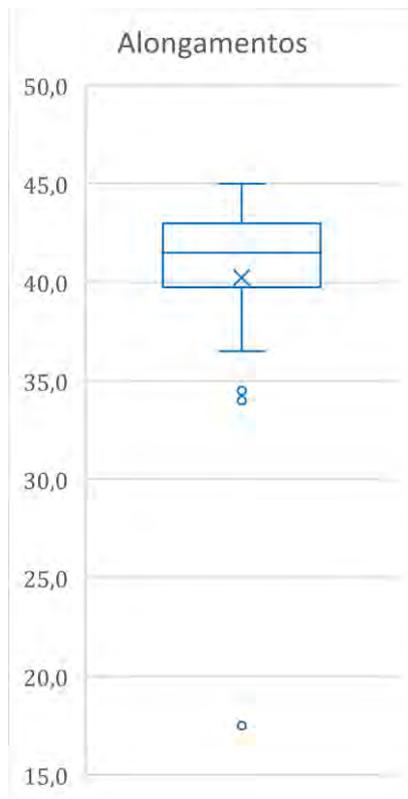
Foram calculadas algumas medidas de localização e dispersão dos dados, que se apresentam a seguir:

Mín	17,5
Máx	45,0
Amplitude	27,5
Média	40,2
Mediana	41,5
Q1	40,3
Q3	43,0
Q3-Q1	2,8
Desvio padrão	5,3

O facto de a média ser inferior à mediana, indica algum enviesamento para a esquerda na distribuição dos dados. Esse enviesamento é visível no seguinte histograma, obtido com o Excel:

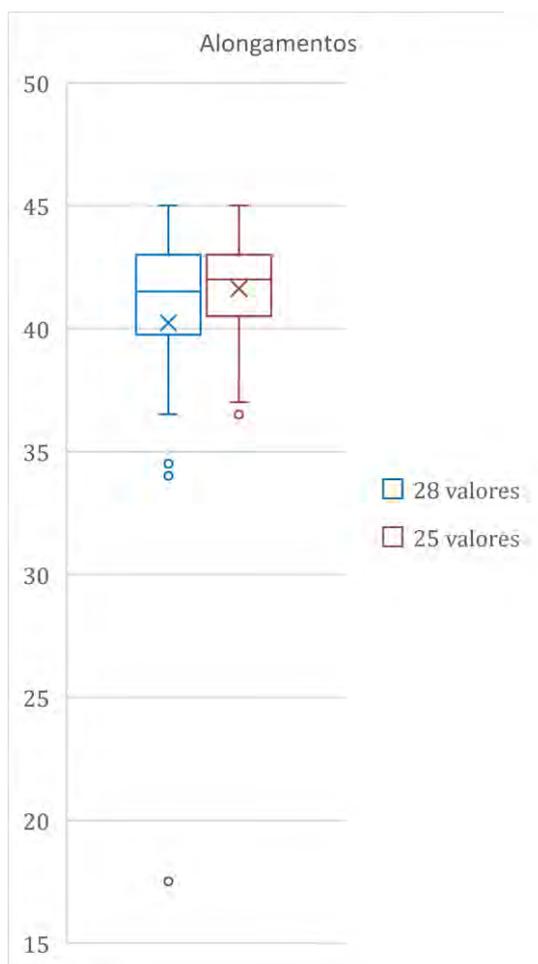


O enviesamento apresentado no histograma, com uma classe com um único elemento e afastada das restantes classes, levou-nos a questionar se não estaríamos perante a presença de um *outlier* (dado cujo valor é bastante diferente dos valores dos restantes dados). A resposta a esta questão foi-nos dada pela representação dos dados num diagrama em caixa de bigodes, também obtido com o Excel:



O diagrama em caixa de bigodes mostra 3 *outliers*, todos eles abaixo do 1º Quartil. São os pontos: 17,5; 34,0; 34,5.

Observámos como se altera o gráfico em caixa de bigodes, ao serem retirados esses três *outliers*. Percebemos que aparece um novo *outlier* no diagrama em caixa de bigodes (o valor 36,5). Esta situação traduz o facto de os valores atípicos serem relativos a cada conjunto de dados que está a ser considerado.



Percebemos também que a assimetria é menor e que há uma menor diferença entre a média e a mediana, como seria de esperar, uma vez que a média não é uma medida resistente, o mesmo acontecendo com o desvio padrão. A média aumentou, passando a ser $\bar{x} = 41,6$ cm. O desvio padrão diminuiu, sendo $s = 2,1$ cm, mostrando agora uma maior concentração em torno da média.

Mín	36,5
Máx	45,0
Amplitude	8,5
Média	41,6
Mediana	42,0
Q1	40,5
Q3	43,0
Q3-Q1	2,5
Desvio padrão	2,1

Assim, constata-se que a variabilidade observada nos elásticos testados é explicada pela existência de *outliers* (neste caso, com valores anormalmente pequenos).

Os *outliers* que surgiram nos dados refletem o facto de os elásticos de borracha perderem, por vezes, as suas propriedades devido a diversos fatores, como temperatura, humidade, oxidação e outros. Com efeito, encontraram-se alguns elásticos “muito frouxos” na amostra estudada (inclusive 2 elásticos quebradiços que se romperam imediatamente) e é razoável admitir que apareçam também noutros pacotes de elásticos do mesmo tipo.

VI. Conclusões e limitações

O estudo incidiu sobre o alongamento máximo de elásticos de secretária numa população composta pelos elásticos de um pacote de 1 kg, de determinado tipo e marca. Foi testada uma amostra de 30 elásticos, dos quais houve 2 que se quebraram de imediato, praticamente sem alongamento. Com os

restantes 28 elásticos estudados ($n = 28$), observou-se grande variabilidade nos alongamentos máximos atingidos. Essa variabilidade foi sobretudo explicada pela existência de *outliers* que se podem classificar como elásticos “muito frouxos”. O alongamento máximo destes elásticos foi anormalmente pequeno, em comparação com a maior concentração de dados da amostra.

Retirando os *outliers* identificados, verificámos que a distribuição ($n = 25$) se tornou menos assimétrica e obtivemos uma média de 41,6 cm e um desvio padrão de 2,1 cm. A amplitude interquartil indica que 50% dos elásticos testados atingiram uma extensão entre 40,5 cm e 43,0 cm. A existência de *outliers* pode ser justificada pelas propriedades da borracha, concretamente pelo facto de esta perder elasticidade em consequência de fatores ambientais.

O processo experimental realizado foi, até certo ponto, semelhante ao que é feito em laboratório, usando máquinas de ensaio próprias. Com essas, é possível obter-se, entre outras medidas, o *Alongamento na Rotura*.

É de notar que os elásticos do tipo Nº 16 têm um comprimento (semiperímetro) de 6 cm, o que significa um comprimento de cerca de 6 cm, entre as pontas, quando em posição plana e de forma reta (como mostra a figura 4).

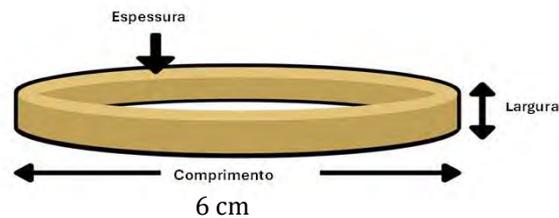


Figura 4. Esquema das dimensões de um elástico

Portanto, sabendo que a média dos alongamentos atingidos foi 41,6 cm (excluindo os *outliers*), conclui-se que o alongamento na rotura foi, em média, segundo a definição: $\frac{41,6-6}{6} = 593\%$. Este resultado é justificado pelo facto de a borracha ser um dos materiais que apresenta alongamentos na rotura mais elevados (Caetano, 2022).

Terminamos com um comentário sobre as limitações deste estudo. É de referir que os testes foram realizados com recurso a um aparato rudimentar; em particular, a força de tensão aplicada pode ter sido diferente de elástico para elástico. Além disso, admite-se que a precisão dos valores obtidos por leitura na régua graduada é baixa.

Apesar das limitações apontadas, o estudo mostrou a importância de se construírem representações gráficas adequadas, na análise dos dados, nomeadamente para identificar dados que possam influenciar os valores das medidas utilizadas para caracterizar os dados. No contexto da situação estudada, os *outliers* têm uma explicação científica relacionada com a deterioração das propriedades mecânicas da borracha (Staughton, 2023). Isto leva-nos a considerar os *outliers* como merecedores de atenção. Os *outliers* podem refletir uma característica do sistema em estudo (ao invés de erros de medição).

VII. Referências

Caetano, M. (2022). *Ciência e Tecnologia da Borracha*. <https://www.ctborracha.com/>

Staughton, J. (2023). *Why do rubber bands lose their elasticity?*. Science ABC. <https://www.scienceabc.com/eyeopeners/why-do-rubber-bands-lose-their-elasticity.html>

COMPREENDER UM ESTUDO SOBRE HIPERTENSÃO NOS JOVENS

TAREFA



Nesta tarefa vamos abordar a associação entre 2 variáveis qualitativas, utilizando tabelas de dupla entrada e percentagens. Ao contrário do que se passa com as variáveis quantitativas, em que se utiliza o coeficiente de correlação para medir a intensidade e tipo de associação, neste caso não temos uma medida equivalente.

Trataremos a associação entre duas variáveis qualitativas tendo por base a leitura e análise de um estudo estatístico realizado por Ricardo Mexia, médico de Saúde Pública, na sua dissertação de mestrado²³.

1. Considera os objetivos que foram explicitados no referido estudo estatístico (p.8):

- 1. Determinar a prevalência de hipertensão arterial na população escolarizada que frequentou o 7º Ano de Escolaridade nas escolas situadas na área de intervenção do Centro de Saúde de Queluz, em 2007.*
- 2. Caracterizar a idade, sexo, perfil estato-ponderal, os antecedentes familiares de hipertensão arterial e a prática de exercício físico da população escolarizada que frequentou o 7º Ano de Escolaridade nas escolas situadas na área de intervenção do Centro de Saúde de Queluz, em 2007, comparando as populações hipertensa e não hipertensa.*

Tendo em conta a informação dada, explica qual era o problema estatístico do estudo. Indica também qual foi a população em estudo.

²³ Mexia, R. (2013). *Hipertensão Arterial em Adolescentes Escolarizados da Área de Intervenção do Centro de Saúde de Queluz*. (Dissertação de Mestrado em Saúde Pública). Escola Nacional de Saúde Pública, Universidade Nova de Lisboa. Disponível em: <https://run.unl.pt/handle/10362/11541>

2. No referido estudo, foram consideradas diversas variáveis, entre as quais, as seguintes (pp. 11-13):

- Tensão arterial (em mmHg)
- Sexo (masculino/feminino)
- IMC (Índice de Massa Corporal) calculado através da fórmula $\frac{\text{Peso (Kg)}}{(\text{Altura (m)})^2}$

Relativamente à variável Tensão Arterial (sistólica e diastólica), foram obtidos os seguintes resultados relativos a 902 indivíduos avaliados:

Tensão arterial sistólica (mm Hg)		Tensão arterial diastólica (mm Hg)	
Mínimo	68	Mínimo	37
Máximo	175	Máximo	120
Média	114	Média	68
Desvio padrão	13,7	Desvio padrão	11,5

Parece-te que a informação dada por estas medidas estatísticas é suficiente para conhecer a prevalência de hipertensão arterial nos adolescentes observados? Explica a tua resposta.

3. As normas estabelecidas pela Direção-Geral da Saúde adotadas no estudo permitiram categorizar os dados da tensão arterial, de tal modo que a variável Tensão Arterial (TA) passou a ser tratada como uma variável qualitativa. Consideraram-se as seguintes categorias²⁴:

- **Tensão Arterial Normal:** TA sistólica e diastólica inferiores ao percentil 90 para a idade e sexo.
- **Tensão Arterial Normal-Alta:** TA sistólica ou diastólica entre os percentis 90 e 95 para a idade e sexo.
- **Hipertensão Arterial:** TA sistólica ou diastólica superior ou igual ao percentil 95 para a idade e sexo.

Recorre às tabelas de percentis utilizadas no estudo²⁵ (percentis de estatura e tensão arterial por percentis de estatura), que encontras em anexo, para categorizar os seguintes casos hipotéticos:

²⁴ Direção Geral da Saúde - DGS. (2005). *Saúde Infantil e Juvenil. Programa-Tipo de Actuação - Orientações Técnicas 12.* (2ª edição). Direção-Geral da Saúde. Divisão de Saúde Materna, Infantil e dos Adolescentes. Disponível em: https://www.mgfamiliar.net/wp-content/uploads/saudeinfantilejuvenil_programatipoactuacao-1.pdf

²⁵ Em 2013 foi publicada uma norma que substituiu o Programa-Tipo de Atuação em Saúde Infantil e Juvenil, de 2005 (ver: <https://www.dgs.pt/directrizes-da-dgs/normas-e-circulares-normativas/norma-n-0102013-de-31052013-jpg.aspx>). O estudo em análise usou a norma anterior.

- i) Rapariga de 14 anos, altura 158 cm, TA 132-75 mmHg (sistólica-diastólica)
- ii) Rapaz de 12 anos, altura 153 cm, TA 110-66 mmHg (sistólica-diastólica)

4. Usando a classificação dos indivíduos por grupos de Tensão Arterial, o estudo apresenta várias tabelas de dupla entrada, nas quais se relacionam duas variáveis qualitativas, usando percentagens. Considera as seguintes tabelas incluídas no estudo (p. 20 e p. 23):

Quadro 8 – Caracterização da tensão arterial, por sexo

Sexo	Tensão Arterial						Total	
	Normal		Normal-Alta		Hipertensão arterial			
	N	%	N	%	N	%	N	%
Feminino	256	46,4%	58	12,8%	140	30,8%	454	50,3%
Masculino	289	64,5%	71	15,9%	88	19,6%	448	49,7%
Total	545	60,4%	129	14,3%	228	25,3%	902	100,0%

Quadro 15 – Caracterização da tensão arterial, por IMC

IMC	Tensão arterial						Total	
	Normal		Normal-Alta		Hipertensão arterial			
	N	%	N	%	N	%	N	%
Elevado	53	41,4%	11	8,6%	64	50,0%	128	14,2%
Normal ou baixo	492	63,6%	118	15,3%	164	21,2%	774	85,8%
Total	545	60,4%	129	14,3%	228	25,3%	902	100,0%

Representa graficamente os resultados que são apresentados nas duas tabelas. Explica a escolha que fizeste do(s) tipo(s) de gráfico(s) usados. Faz a sua interpretação.

- 5. Considerando que a hipertensão e até a tensão normal-alta são preocupantes, obtém a percentagem de casos com tensão arterial preocupante: nas raparigas; nos rapazes; nos indivíduos com IMC normal ou baixo; nos indivíduos com IMC elevado.
- 6. Por comparação, qual te parece ser o fator de maior risco de hipertensão arterial na população observada: ser do sexo feminino ou ter um elevado IMC? Justifica.
- 7. Nas conclusões do estudo, é afirmado o seguinte (p. 31):

Em conclusão, este estudo permitiu demonstrar que a HTA tem uma prevalência elevada na população adolescente estudada, atingindo valores que alguns estudos já tinham registado. Isto é particularmente preocupante, se atendermos também ao facto de que o valor da tensão normal-alta somado ao da hipertensão atinge cerca de 40%.

Explica de que forma a conclusão obtida está apoiada nos resultados estatísticos.

ANEXO

Gráfico 1 – Estatura das raparigas dos 2 aos 20 anos

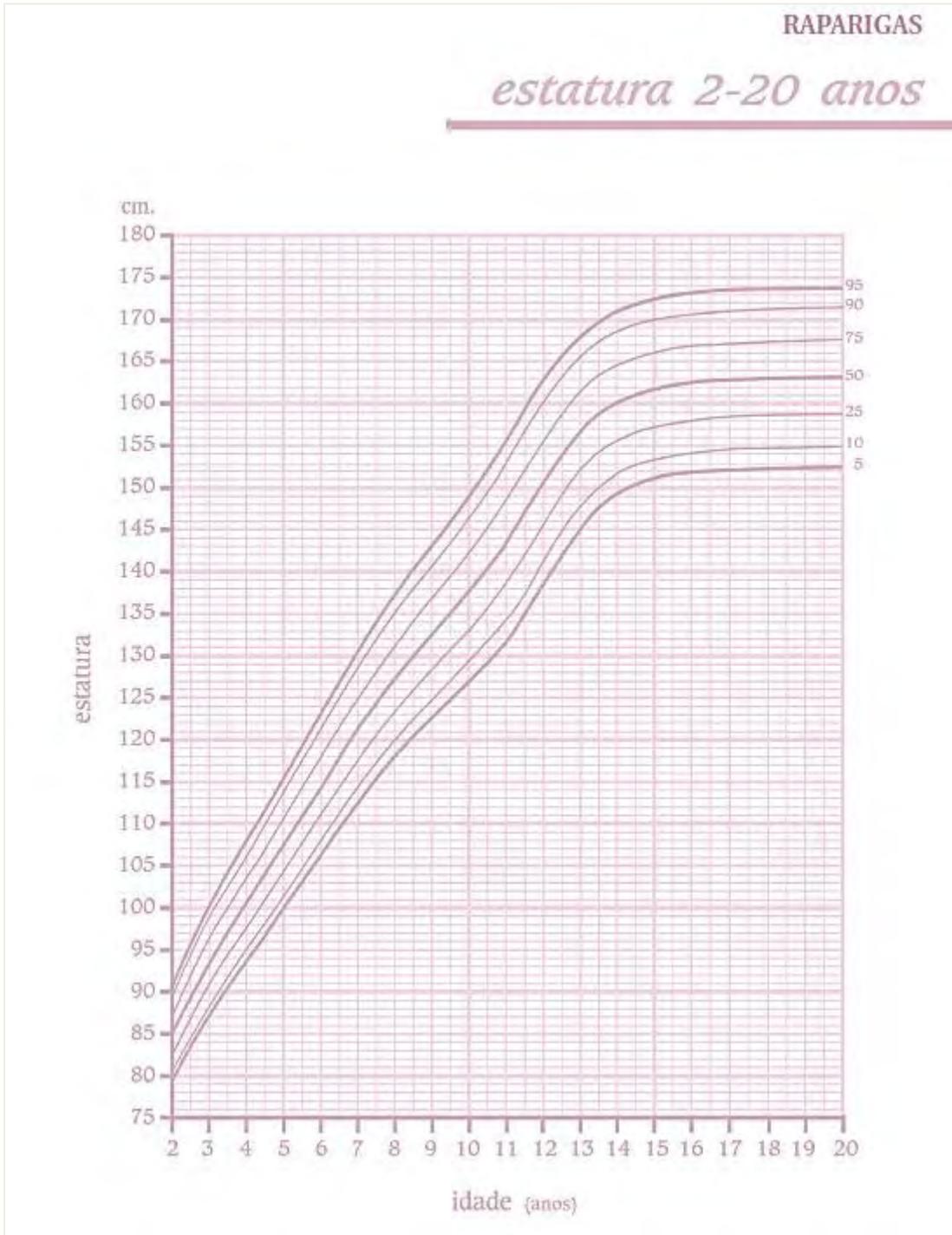


Tabela 1 – Valores de Tensão Arterial Sistólica por Percentis de Estatura (raparigas dos 1 aos 17 anos)

TABELA 1
Valores de Tensão Arterial SISTÓLICA por Percentis de Estatura
RAPARIGAS 1 A 17 ANOS

Idade (anos)	Percentil Tensão arterial *	Tensão arterial sistólica / percentil estatura mm Hg**						
		5%	10%	25%	50%	75%	90%	95%
1	90	97	98	99	100	102	103	104
	95	101	102	103	104	105	107	107
2	90	99	99	100	102	103	104	105
	95	102	103	104	105	107	108	109
3	90	100	100	102	103	104	105	106
	95	104	104	105	107	108	109	110
4	90	101	102	103	104	106	107	108
	95	105	106	107	108	109	111	111
5	90	103	103	104	106	107	108	109
	95	107	107	108	110	111	112	113
6	90	104	105	106	107	109	110	111
	95	108	109	110	111	112	114	114
7	90	106	107	108	109	110	112	112
	95	110	110	112	113	114	115	116
8	90	108	109	110	111	112	113	114
	95	112	112	113	115	116	117	118
9	90	110	110	112	113	114	115	116
	95	114	114	115	117	118	119	120
10	90	112	112	114	115	116	117	118
	95	116	116	117	119	120	121	122
11	90	114	114	116	117	118	119	120
	95	118	118	119	121	122	123	124
12	90	116	116	118	119	120	121	122
	95	120	120	121	123	124	125	126
13	90	118	118	119	121	122	123	124
	95	121	122	123	125	126	127	128
14	90	119	120	121	122	124	125	126
	95	123	124	125	126	128	129	130
15	90	121	121	122	124	125	126	127
	95	124	125	126	128	129	130	131
16	90	122	122	123	125	126	127	128
	95	125	126	127	128	130	131	132
17	90	122	123	124	125	126	128	128
	95	126	126	127	129	130	131	132

* Percentil de tensão arterial determinada por uma única leitura
** Percentil de estatura determinado nas curvas-padrão de crescimento

Tabela 2 – Valores de Tensão Arterial Diastólica por Percentis de Estatura (raparigas dos 1 aos 17 anos)

TABELA 2
Valores de Tensão Arterial DIASTÓLICA por Percentis de Estatura
RAPARIGAS 1 A 17 ANOS

Idade (anos)	Percentil Tensão arterial *	Tensão arterial diastólica / percentil estatura mm Hg**						
		5%	10%	25%	50%	75%	90%	95%
1	90	53	53	53	54	55	56	56
	95	57	57	57	58	59	60	60
2	90	57	57	58	58	59	60	61
	95	61	61	62	62	63	64	65
3	90	61	61	61	62	63	63	64
	95	65	65	65	66	67	67	68
4	90	63	63	64	65	65	66	67
	95	67	67	68	69	69	70	71
5	90	65	66	66	67	68	68	69
	95	69	70	70	71	72	72	73
6	90	67	67	68	69	69	70	71
	95	71	71	72	73	73	74	75
7	90	69	69	69	70	71	72	72
	95	73	73	73	74	75	76	76
8	90	70	70	71	71	72	73	74
	95	74	74	75	75	76	77	78
9	90	71	72	72	73	74	74	75
	95	75	76	76	77	78	78	79
10	90	73	73	73	74	75	76	76
	95	77	77	77	78	79	80	80
11	90	74	74	75	75	76	77	77
	95	78	78	79	79	80	81	81
12	90	75	75	76	76	77	78	78
	95	79	79	80	80	81	82	82
13	90	76	76	77	78	78	79	80
	95	80	80	81	82	82	83	84
14	90	77	77	78	79	79	80	81
	95	81	81	82	83	83	84	85
15	90	78	78	79	79	80	81	82
	95	82	82	83	83	84	85	86
16	90	79	79	79	80	81	82	82
	95	83	83	83	84	85	86	86
17	90	79	79	79	80	81	82	82
	95	83	83	83	84	85	86	86

* Percentil de tensão arterial determinada por uma única leitura
** Percentil de estatura determinado nas curvas-padrão de crescimento

Gráfico 2 – Estatura dos rapazes dos 2 aos 20 anos

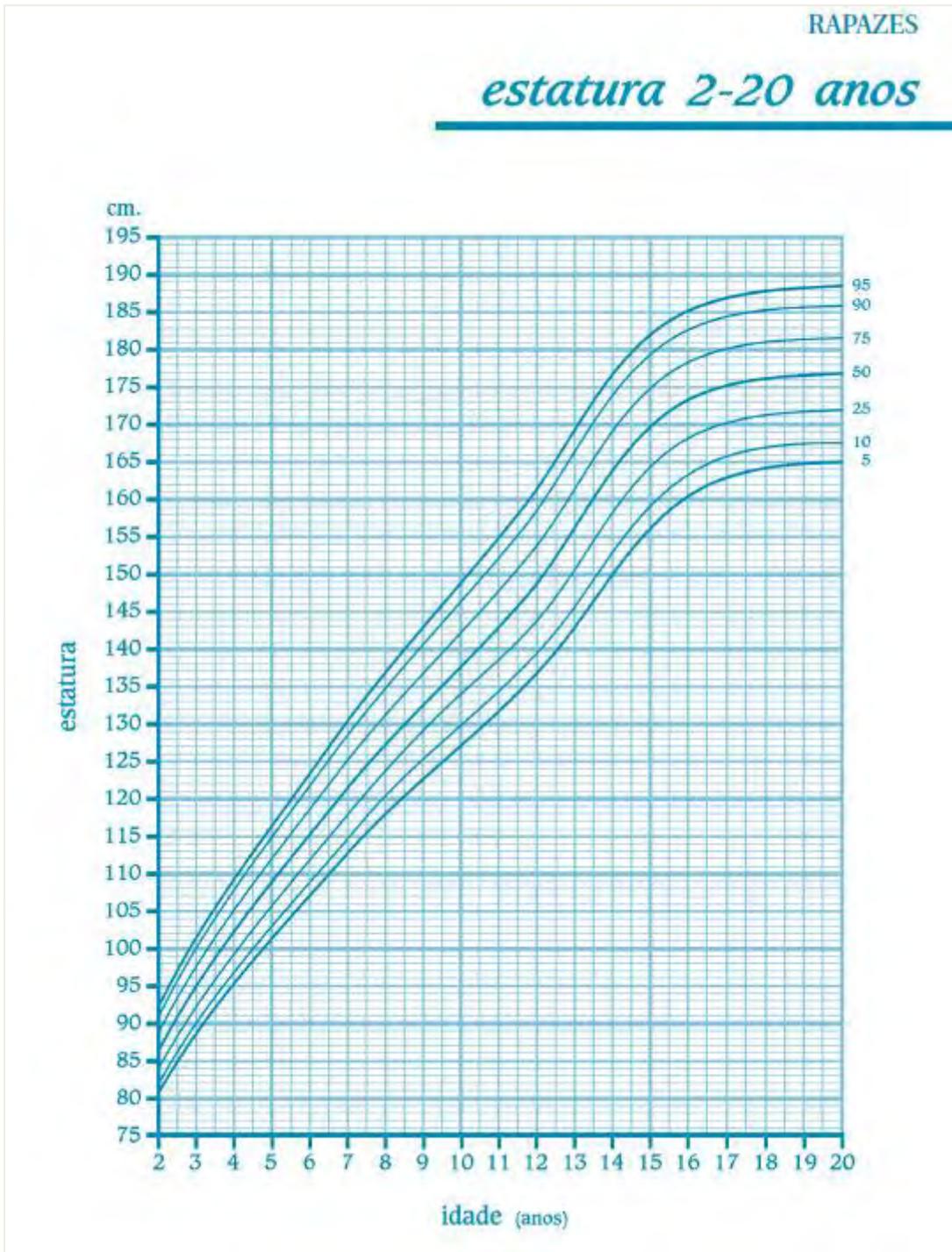


Tabela 3 – Valores de Tensão Arterial Sistólica por Percentis de Estatura (rapazes dos 1 aos 17 anos)

TABELA 3
Valores de Tensão Arterial SISTÓLICA por Percentis de Estatura
RAPAZES 1 A 17 ANOS

Idade (anos)	Percentil Tensão arterial *	Tensão arterial sistólica / percentil estatura mm Hg**						
		5%	10%	25%	50%	75%	90%	95%
1	90	94	95	97	98	100	102	102
	95	98	99	101	102	104	106	106
2	90	98	99	100	102	104	105	106
	95	101	102	104	106	108	109	110
3	90	100	101	103	105	107	108	109
	95	104	105	107	109	111	112	113
4	90	102	103	105	107	109	110	111
	95	106	107	109	111	113	114	115
5	90	104	105	106	108	110	112	112
	95	108	109	110	112	114	115	116
6	90	105	106	108	110	111	113	114
	95	109	110	112	114	115	117	117
7	90	106	107	109	111	113	114	115
	95	110	111	113	115	116	118	119
8	90	107	108	110	112	114	115	116
	95	111	112	114	116	118	119	120
9	90	109	110	112	113	115	117	117
	95	113	114	116	117	119	121	121
10	90	110	112	113	115	117	118	119
	95	114	115	117	119	121	122	123
11	90	112	113	115	117	119	120	121
	95	116	117	119	121	123	124	125
12	90	115	116	117	119	121	123	123
	95	119	120	121	123	125	126	127
13	90	117	118	120	122	124	125	126
	95	121	122	124	126	128	129	130
14	90	120	121	123	125	126	128	128
	95	124	125	127	128	130	132	132
15	90	123	124	125	127	129	131	131
	95	127	128	129	131	133	134	135
16	90	125	126	128	130	132	133	134
	95	129	130	132	134	136	137	138
17	90	128	129	131	133	134	136	136
	95	132	133	135	136	138	140	140

* Percentil de tensão arterial determinada por uma única leitura
** Percentil de estatura determinado nas curvas-padrão de crescimento

Tabela 4 – Valores de Tensão Arterial Diastólica por Percentis de Estatura (rapazes dos 1 aos 17 anos)

TABELA 4
Valores de Tensão Arterial DIASTÓLICA por Percentis de Estatura
RAPAZES 1 A 17 ANOS

Idade (anos)	Percentil Tensão arterial *	Tensão arterial diastólica / percentil estatura mm Hg**						
		5%	10%	25%	50%	75%	90%	95%
1	90	50	51	52	53	54	54	55
	95	55	55	56	57	58	59	59
2	90	55	55	56	57	58	59	59
	95	59	59	60	61	62	63	63
3	90	59	59	60	61	62	63	63
	95	63	63	64	65	66	67	67
4	90	62	62	63	64	65	66	66
	95	66	67	67	68	69	70	71
5	90	65	65	66	67	68	69	69
	95	69	70	70	71	72	73	74
6	90	67	68	69	70	70	71	72
	95	72	72	73	74	75	76	76
7	90	69	70	71	72	72	73	74
	95	74	74	75	76	77	78	78
8	90	71	71	72	73	74	75	75
	95	75	76	76	77	78	79	80
9	90	72	73	73	74	75	76	77
	95	76	77	78	79	80	80	81
10	90	73	74	74	75	76	77	78
	95	77	78	79	80	80	81	82
11	90	74	74	75	76	77	78	78
	95	78	79	79	80	81	82	83
12	90	75	75	76	77	78	78	79
	95	79	79	80	81	82	83	83
13	90	75	76	76	77	78	79	80
	95	79	80	81	82	83	83	84
14	90	76	76	77	78	79	80	80
	95	80	81	81	82	83	84	85
15	90	77	77	78	79	80	81	81
	95	81	82	83	83	84	85	86
16	90	79	79	80	81	82	82	83
	95	83	83	84	85	86	87	87
17	90	81	81	82	83	84	85	85
	95	85	85	86	87	88	89	89

* Percentil de tensão arterial determinada por uma única leitura
** Percentil de estatura determinado nas curvas-padrão de crescimento

NOTAS PARA O PROFESSOR

A exploração da tarefa **Compreender um estudo sobre hipertensão nos jovens** é adequada às Aprendizagens Essenciais de Matemática A, Matemática B, MACS e Matemática dos Cursos Profissionais, como um possível aprofundamento dos temas estudados.

A tarefa tem um grande foco no desenvolvimento da Educação para a Saúde, incidindo sobre a problemática da tensão arterial na população mais jovem, permitindo consciencializar os alunos para o problema da hipertensão em Portugal, que afeta muitos adolescentes, promovendo a autonomia dos alunos, o trabalho interdisciplinar, a comunicação matemática e o desenvolvimento do espírito crítico.

Nesta tarefa trata-se a associação entre duas variáveis qualitativas, tendo por base a leitura e análise de um estudo estatístico realizado por Ricardo Mexia, médico de Saúde Pública, na sua dissertação de mestrado. A resolução da tarefa, a partir dos dados disponíveis no estudo, pretende levar os alunos a refletir sobre a importância de estudos científicos para o avanço do conhecimento e familiarizar os alunos com a análise e interpretação de dados e resultados estatísticos disponíveis nos mesmos.

Tema: Estatística

Tópicos e subtópicos:

População, amostra e variável

- Fases de um procedimento estatístico

Dados univariados

-Medidas de localização

- Medidas de dispersão

- Propriedades das medidas

Objetivos de aprendizagem:

- Identificar num estudo estatístico, população, amostra e a(s) característica(s) a estudar, que se designa(m) por variável (variáveis).
- Reconhecer as fases de um procedimento estatístico: i) produção ou aquisição de dados; ii) organização e representação de dados e iii) interpretação tendo por base as representações obtidas.
- Interpretar as medidas de localização: média (\bar{x}) e percentis na caracterização da distribuição dos dados.
- Interpretar as medidas de dispersão, amplitude e desvio padrão amostral, s , na caracterização da distribuição dos dados,
- Compreender que uma das propriedades das medidas é a pouca resistência da média e do desvio padrão.

A tarefa poderá ser tratada como possível aprofundamento dos conteúdos trabalhados no 10.º ano de escolaridade, uma vez que aborda a associação entre variáveis qualitativas ou categóricas.

Duração prevista: Duas aulas de 50 minutos.

Materiais e recursos:

- Calculadora gráfica, computador.
- Internet, folha de cálculo.

Estratégias de implementação:

Esta tarefa aborda a associação entre 2 variáveis qualitativas, utilizando tabelas de dupla entrada e percentagens. Ao contrário do que se passa com as variáveis quantitativas, em que se utiliza o coeficiente de correlação para medir a intensidade e tipo de associação, neste caso não temos uma medida equivalente.

Simultaneamente, pretende-se evidenciar a importância dos estudos científicos para o conhecimento de problemas que afetam a nossa sociedade, em particular na área da saúde, e sensibilizar os alunos para o perigo da hipertensão arterial, com enfoque nas idades jovens.

Recomendamos que a tarefa possa ser resolvida em grupos, para facilitar a análise dos dados disponibilizados e promover a discussão entre pares.

No final, poderá ser promovida uma discussão coletiva sobre o tema estudado, pedindo-se a um ou mais grupos que apresentem as suas conclusões e resolvendo no quadro a tarefa. Finalmente, a partir das resoluções apresentadas poderá ser feita, de forma colaborativa, uma síntese das principais ideias trabalhadas ao longo da resolução da tarefa, destacando a inexistência de um coeficiente que permita relacionar pares de variáveis qualitativas. No final da aula, os alunos refletem individualmente sobre o seu trabalho e o que aprenderam.

Dificuldades esperadas e ações do professor:

Os alunos podem revelar algumas dificuldades, nomeadamente em:

- Analisar e compreender os dados disponibilizados e trabalhar com muita informação.
- Compreender que não poderão usar, na análise de dados qualitativos, ferramentas de análise estatística que são usadas em dados bivariados quantitativos.
- Fazer a leitura cruzada de gráficos e tabelas de percentis e usar os critérios adotados para a classificação dos dados.

O professor, deverá acompanhar o trabalho dos diferentes grupos, solicitando aos mesmos que verbalizem a interpretação dos dados, dando sugestões ou colocando questões que contribuam para o desenvolvimento do trabalho.

Avaliação:

A avaliação poderá ter um carácter formativo. Durante a aula o professor poderá dar aos alunos feedback oral sobre o seu desempenho na resolução da tarefa.

PROPOSTA DE RESOLUÇÃO

1. O problema estatístico consistiu em conhecer a prevalência da hipertensão arterial numa população de adolescentes, num determinado momento, bem como perceber se a hipertensão poderia estar associada a certas variáveis. A prevalência significa o número de casos numa dada população e exprime-se geralmente sob a forma de percentagem.

A população em estudo foi constituída pelos indivíduos que, em 2007, frequentavam o 7^o ano de escolaridade nas escolas situadas na área de intervenção do Centro de Saúde de Queluz.

2. Apesar de termos algumas medidas de localização e de dispersão da tensão arterial sistólica e da tensão arterial diastólica, não temos informação suficiente para descrever a distribuição de cada uma dessas variáveis. Note-se, por exemplo, que a média e o desvio padrão são medidas sensíveis à existência de valores atípicos.

Por outro lado, para se determinar a prevalência da hipertensão arterial, terá de se definir o que se entende por hipertensão, que é diferente para o caso dos adultos e para o caso das crianças ou adolescentes. Segundo informação disponível no site *Pedipedia*²⁶, a hipertensão arterial em crianças e adolescentes é definida estatisticamente, através de percentis, como se segue:

“Ao contrário do que acontece nos adultos, a hipertensão arterial pediátrica é uma definição estatística baseada na distribuição da pressão arterial em crianças e adolescentes saudáveis, sendo que os valores da mesma aumentam com a idade e o tamanho corporal. Assim a distribuição destes valores permitiu o estabelecimento de percentis, que variam com a idade, género e estatura”.

3. Em primeiro lugar será determinado, aproximadamente, o percentil da estatura conforme o sexo, com base no gráfico de percentis. A seguir, nas tabelas da tensão arterial (sistólica e diastólica) relativas ao sexo do indivíduo, faz-se o cruzamento entre a idade e o percentil da estatura para obter os valores dos **percentis 90 e 95 da tensão arterial**.

i) Rapariga de 14 anos, altura 158 cm, TA 132-75 mmHg

Através da leitura do gráfico de percentis de estatura em raparigas, foi obtido, aproximadamente, o percentil 50 para a estatura.

Na tabela da tensão arterial **sistólica**, para raparigas, na linha relativa à idade de 14 anos e na coluna do percentil 50 de estatura, encontraram-se os valores dos percentis 90 e 95. Da mesma forma, na tabela da tensão arterial **diastólica**, para raparigas, encontraram-se os percentis 90 e 95. Os valores são os seguintes:

Tensão arterial sistólica	Percentil 90	122
	Percentil 95	126
Tensão arterial diastólica	Percentil 90	79
	Percentil 95	83

Com estes valores, de acordo com os critérios indicados, a categorização será: Hipertensão Arterial (porque a TA sistólica é superior ou igual ao percentil 95 para a idade e sexo).

ii) Rapaz de 12 anos, altura 153 cm, TA 110-66 mm Hg

²⁶ Ver: <https://pedipedia.org/artigo-profissional/hipertensao-arterial-na-crianca-e-no-adolescente>

Através da leitura do gráfico de percentis de estatura em rapazes, foi obtido, aproximadamente, o percentil 75 para a estatura.

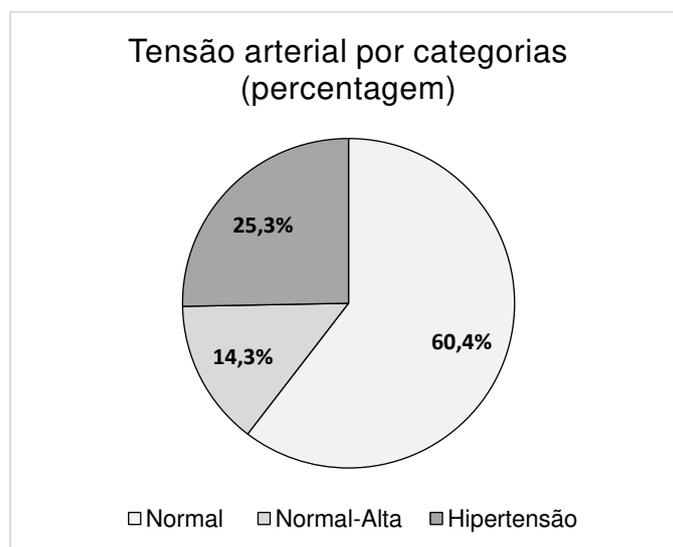
Na tabela da tensão arterial **sistólica**, para rapazes, na linha relativa à idade de 12 anos e na coluna do percentil 75 de estatura, encontraram-se os valores dos percentis 90 e 95. Da mesma forma, na tabela da tensão arterial **diastólica**, para rapazes, encontraram-se os percentis 90 e 95. Os valores são os seguintes:

Tensão arterial sistólica	Percentil 90	121
	Percentil 95	125
Tensão arterial diastólica	Percentil 90	78
	Percentil 95	82

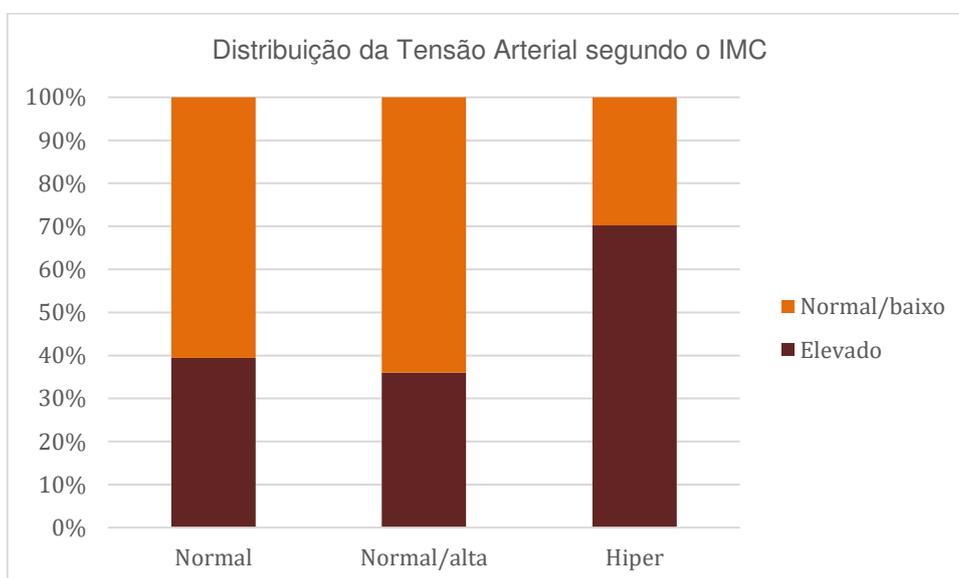
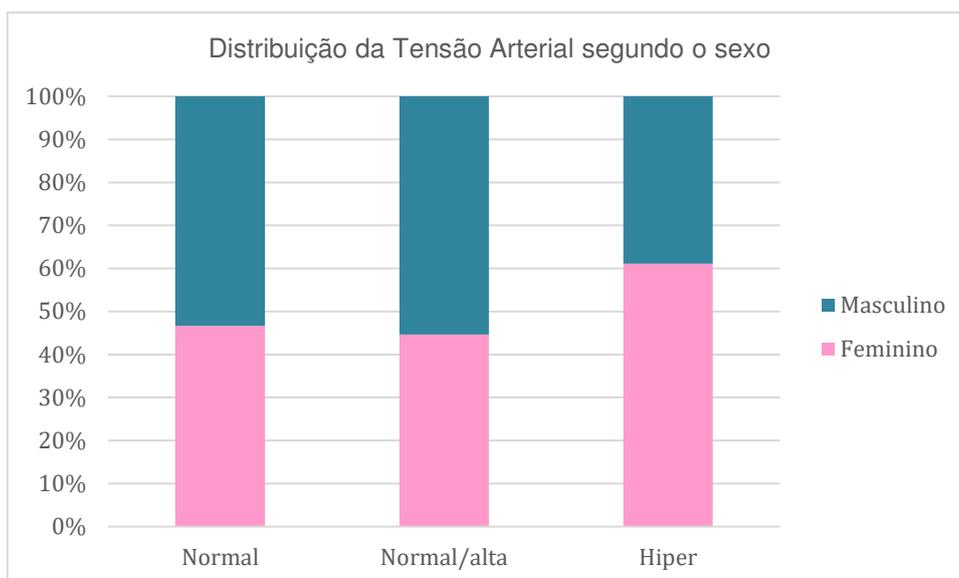
Com estes valores, de acordo com os critérios indicados, a categorização será: Tensão Arterial Normal (porque a TA sistólica e a TA diastólica são ambas inferiores ao percentil 90 para a idade e sexo).

4. Seguem-se os gráficos construídos na Folha de Cálculo para representar os dados das duas tabelas.

O primeiro gráfico mostra a percentagem de indivíduos de cada uma das três categorias de tensão arterial. Foi escolhido o gráfico circular pois é útil para representar as percentagens de cada categoria de uma mesma variável, uma vez que são poucas categorias. Com este tipo de gráfico, pode-se comparar rapidamente cada um dos setores com o total. Dado que o estudo pretende analisar a prevalência da hipertensão na população estudada, o gráfico circular torna claro que essa prevalência é cerca de 25%, ou seja, 1 em 4 adolescentes é hipertenso.



Os dois gráficos seguintes mostram a percentagem de cada uma das categorias de tensão arterial, por sexo e por IMC, respetivamente. Foi escolhido o gráfico de colunas empilhadas a 100%, porque mostra as frequências relativas de uma variável qualitativa, segundo outra variável qualitativa. Este tipo de gráfico permite facilmente visualizar subcategorias como, por exemplo, a subcategoria de “indivíduos com hipertensão e do sexo masculino”.



Observa-se que a percentagem de hipertensão é maior no sexo feminino do que no sexo masculino. Também a percentagem de hipertensão é maior nos adolescentes que têm IMC elevado do que naqueles que têm IMC normal ou baixo.

5. Percentagens de adolescentes com hipertensão ou tensão normal-alta (considerados casos preocupantes), por sexo e por nível de IMC:

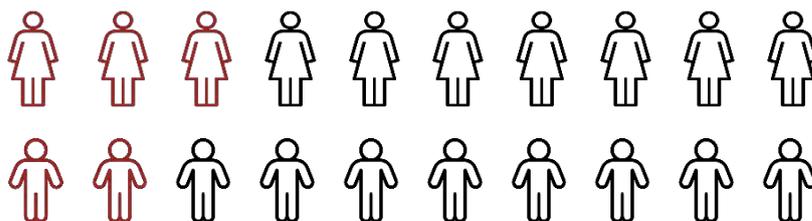
Do sexo feminino: 43,6% (30,8% + 12,8%)

Do sexo masculino: 35,5% (19,6% + 15,9%)

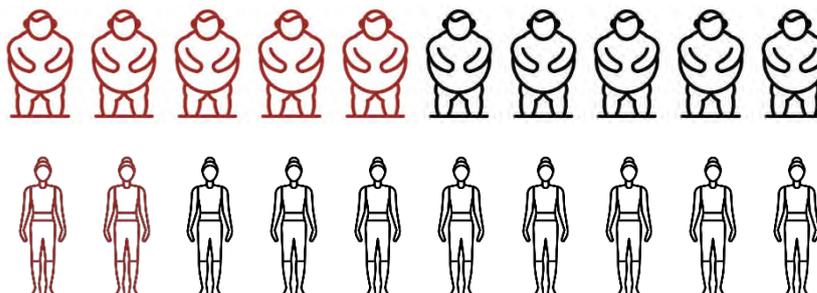
Com IMC elevado: 58,6% (50,0% + 8,6%)

Com IMC normal ou baixo: 36,5% (21,2% + 15,3%)

6. A prevalência de hipertensão nas raparigas é de cerca de 30% (3 em 10 têm hipertensão); no caso dos rapazes, a prevalência de hipertensão é de cerca de 20% (2 em 10 têm hipertensão). A figura seguinte ilustra esta situação.



A mesma análise feita para a variável IMC mostra que a prevalência de hipertensão nos indivíduos com IMC elevado é cerca de 50% (5 em 10 têm hipertensão) e a prevalência de hipertensão em indivíduos com IMC normal-baixo é cerca de 20% (2 em 10 têm hipertensão). A figura seguinte ilustra essa conclusão.



Por outro lado, se analisarmos apenas o grupo dos adolescentes hipertensos, percebemos que o número de indivíduos do sexo feminino é superior ao de indivíduos do sexo masculino; tal proporção é aproximadamente de 3:2.

Da mesma forma, pode ser analisada, no grupo dos hipertensos, a proporção entre casos com IMC elevado e casos com IMC normal ou baixo; tal proporção é aproximadamente de 5:2.

Assim, com base nos dados, e para a população considerada, ter um IMC elevado significa uma maior probabilidade de ter hipertensão do que ser do sexo feminino.

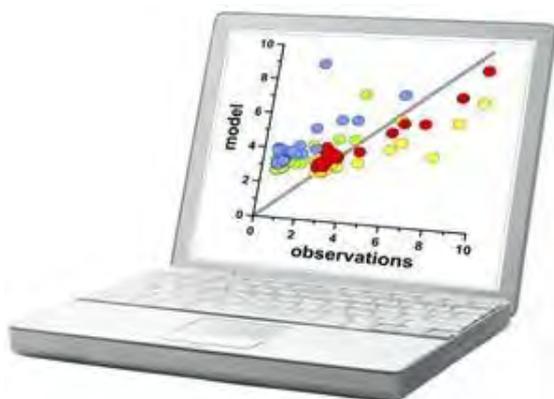
6. Uma das conclusões do estudo foi a seguinte:

Em conclusão, este estudo permitiu demonstrar que a HTA tem uma prevalência elevada na população adolescente estudada, atingindo valores que alguns estudos já tinham registado. Isto é particularmente preocupante, se atendermos também ao facto de que o valor da tensão normal-alta somado ao da hipertensão atinge cerca de 40%.

A conclusão expõe a prevalência de hipertensão na população de adolescentes estudada, que era um dos dois objetivos do estudo. De acordo com os resultados estatísticos, essa prevalência foi de 25,3%. A conclusão salienta ainda que a soma das percentagens de adolescentes com hipertensão e com tensão normal-alta ($25,3\% + 14,3\% = 39,6\%$) é próxima de 40%. A prevalência de hipertensão na população estudada foi interpretada pelo autor do estudo como elevada e preocupante, considerando também resultados de estudos anteriores. A conclusão formulada é apoiada pelos resultados estatísticos do estudo.

O COEFICIENTE DE CORRELAÇÃO LINEAR

PARA LER E COMPREENDER



Para analisar a associação entre variáveis quantitativas, através de um conjunto de dados bivariados, o primeiro passo consiste na visualização prévia desses dados num *diagrama de dispersão* e considerar o aspeto da *nuvem de pontos* que nos fornece informação sobre a **forma, direção e grau de associação** das variáveis.

Se se concluir que tem sentido falar numa associação linear entre as variáveis, traduzida pela nuvem de pontos com a “forma de uma oval”, “mais ou menos alongada”, então passa-se a uma fase posterior, que será a medição do “grau de intensidade com que as variáveis se associam”, utilizando o **coeficiente de correlação linear** ou a construção de um modelo que permita conhecer como se refletem numa das variáveis, as modificações processadas na outra, que são os **modelos de regressão**. Aqui, iremos trabalhar unicamente sobre o processo de medir a intensidade com que as variáveis se associam linearmente.

Coeficiente de correlação linear

A medida que se utiliza com mais frequência para medir o grau da associação linear, é o coeficiente de correlação, que se representa por r , e se calcula a partir da expressão:

$$r = \frac{S_{xy}}{\sqrt{S_{xx}} \sqrt{S_{yy}}}$$

onde:

$S_{xy} = \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})$ em que (x_i, y_i) , com $i = 1, \dots, n$, representam os dados bivariados e em que \bar{x} e \bar{y} são as médias dos x_i e y_i , respetivamente.

Propriedades do coeficiente de correlação²⁷:

- 1 - O valor de r está no intervalo $[-1,1]$.
- 2 - Quanto maior for o módulo de r , maior será a relação linear existente entre os x_i e os y_i .
- 3 - O facto de r ser positivo, significa que a relação entre os x_i e os y_i é do mesmo sentido, isto é, a valores grandes de x , correspondem, em média, valores grandes de y e vice-versa. Quando r é negativo, a relação entre os x_i e os y_i é de sentido contrário, o que significa que a valores grandes de x , correspondem, em média, valores pequenos de y e vice-versa.

Interpretação geométrica do coeficiente de correlação

Vejamos agora como se pode interpretar geometricamente o coeficiente de correlação linear. Para isso, vamos considerar as situações A) e B) em que se faz essa interpretação geométrica. Na Parte III da tarefa sugerimos que faça uma interpretação análoga para a situação que te é proposta, usando os dados que são disponibilizados para o efeito.

Situação A):

Se aos maiores valores de x estão associados, de um modo geral, os maiores valores de y , então $r > 0$. Efetivamente, quando pensamos num valor grande de x , será um valor acima da média. Por outro lado, um valor pequeno de x é um valor abaixo da média. Então se, de um modo geral, aos valores grandes de x estão associados os valores grandes de y , e aos valores pequenos de x estão associados os valores pequenos de y , os produtos

$$(x - \bar{x})(y - \bar{y})$$

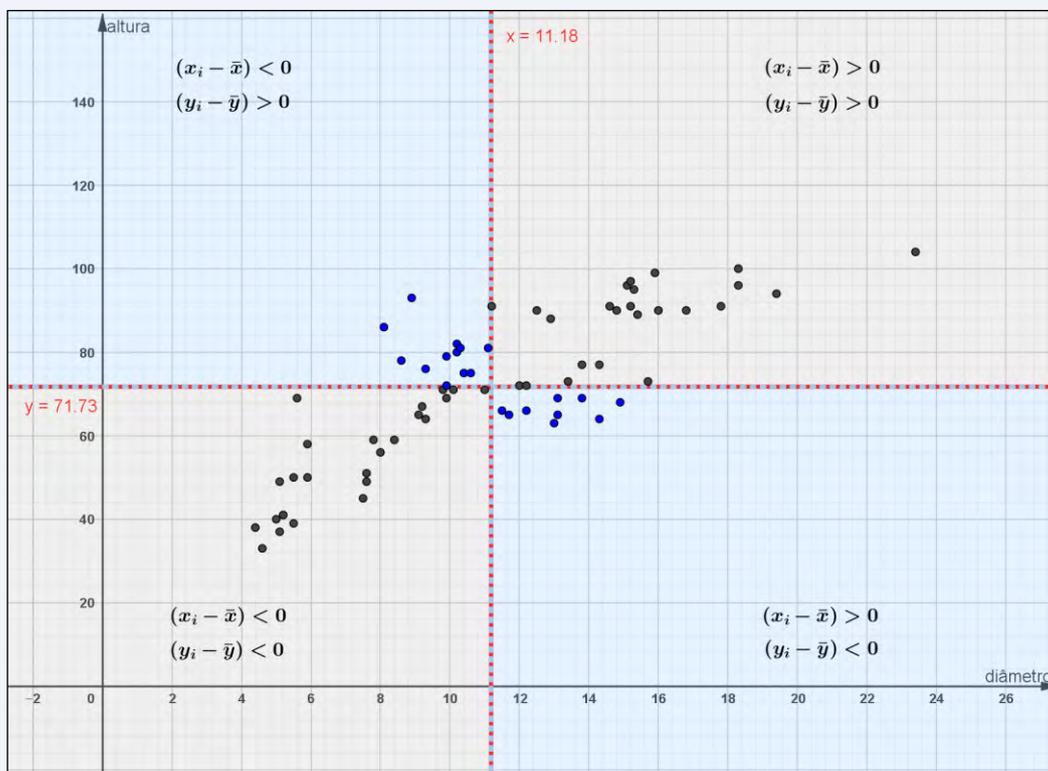
são de um modo geral positivos, já que ambos os fatores são positivos ou negativos. Como o denominador da expressão do coeficiente de correlação, não depende da forma como os valores de x se associam com os valores de y , então o facto de, no numerador, somarmos grande número de parcelas positivas faz com que o valor do coeficiente de correlação seja positivo e tanto maior quantas mais parcelas positivas houver.

Para ilustrarmos geometricamente esta conclusão, analisaremos os dados bivariados relativos ao tamanho de pinheiros²⁸ ($n = 70$), considerando as variáveis: *diâmetro do tronco* (x) e *altura* (y) do pinheiro, expressas em centímetros. Usaremos o GeoGebra para a análise dos dados.

²⁷ Quando se disser coeficiente de correlação, é pressuposto que se trata do coeficiente de correlação linear.

²⁸ Dados disponíveis em: <https://www.statcrunch.com/app/index.html?dataid=57497>

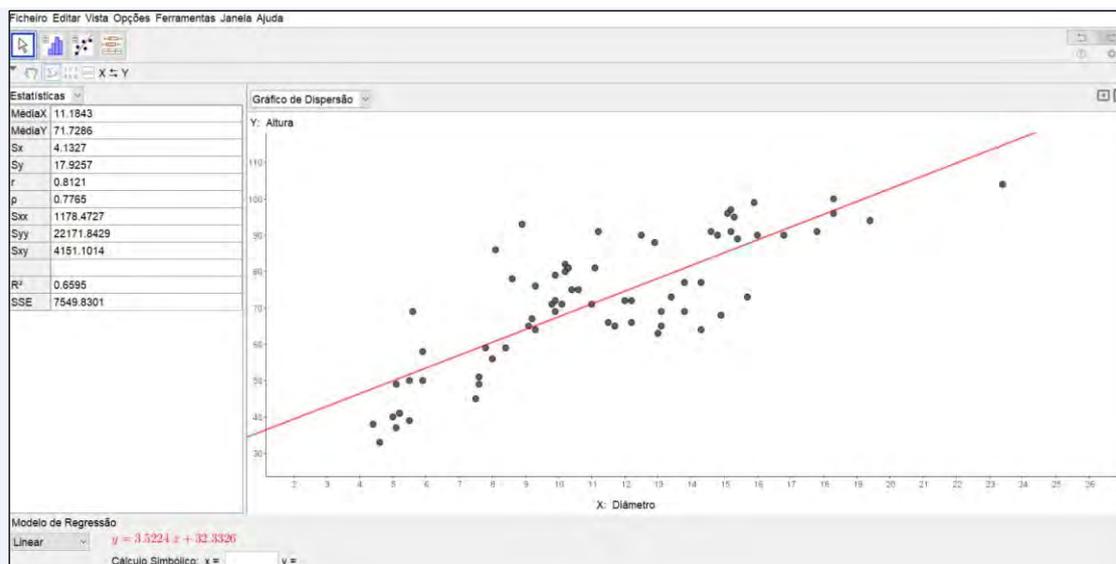
Na janela Folha de Cálculo foram introduzidos os dados, em duas colunas. Foi criada e exibida a lista de pontos no gráfico X-Y (ou gráfico de dispersão). Foram ainda calculadas as médias de cada um dos conjuntos de dados: $\bar{x} = 11,18$ e $\bar{y} = 71,73$. Foi representada a reta vertical $x = 11,18$ e a reta horizontal $y = 71,73$.



No gráfico, é possível visualizar os pontos marcados a preto que se encontram nos dois quadrantes em que os produtos $(x - \bar{x})(y - \bar{y})$ são positivos. Analogamente, estão marcados a azul os pontos que se encontram nos dois quadrantes em que os produtos $(x - \bar{x})(y - \bar{y})$ são negativos.

Pela análise do gráfico, observa-se que o número de pontos a preto é maior do que o número de pontos a azul, ou seja, o número de parcelas positivas é maior do que o número de parcelas negativas.

Ainda no GeoGebra, foi utilizada a ferramenta Análise de Regressão Bivariada, com a qual se obteve a equação da reta de regressão $y = 3,5x + 32,3$ e o valor do coeficiente de correlação, $r = 0,8121$, que significa uma correlação linear positiva, moderada/forte entre as variáveis.



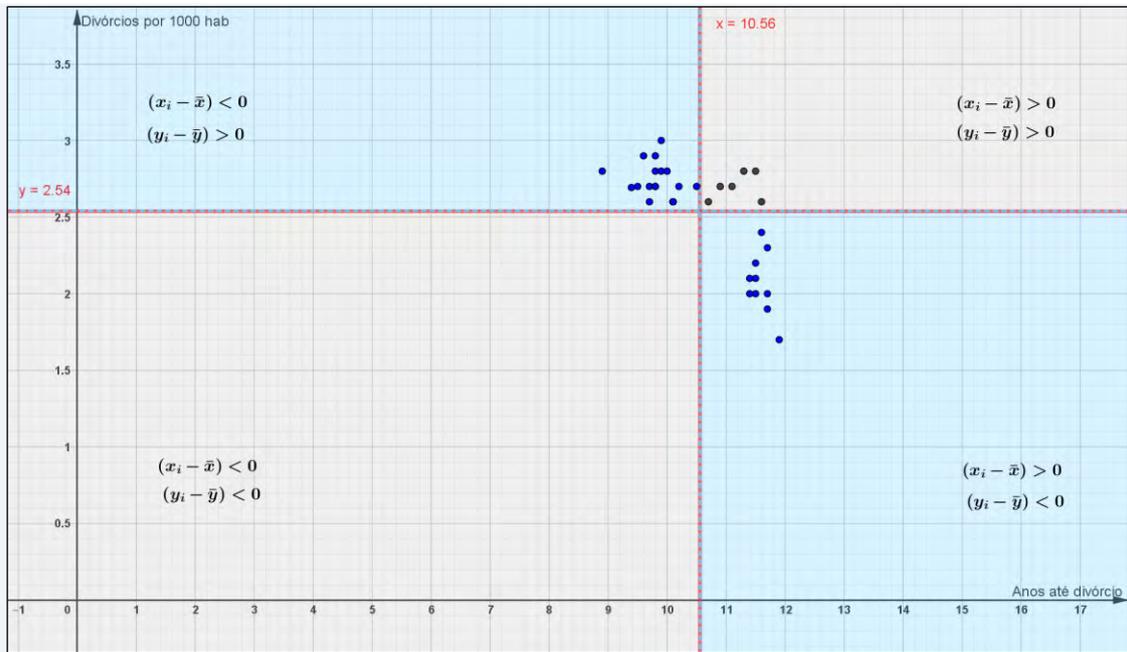
Situação B)

Se aos maiores valores de x estão associados, de um modo geral, os menores valores de y , então $r < 0$. Raciocinando como no exemplo anterior, diremos agora que as parcelas são maioritariamente negativas, já que quando x é grande (superior à média dos valores da variável x), então y é, de um modo geral, pequeno (inferior à média dos valores da variável y). Assim, os produtos $(x - \bar{x})(y - \bar{y})$ são, de um modo geral, negativos.

Neste caso, a ilustração geométrica segue o mesmo processo, agora com dados bivariados relativos aos divórcios ocorridos anualmente no Reino Unido²⁹, entre 1980 e 2015, em que se consideraram as variáveis: *duração média (em anos) dos casamentos até ao divórcio* (x) e *número de divórcios por 1000 habitantes* (y).

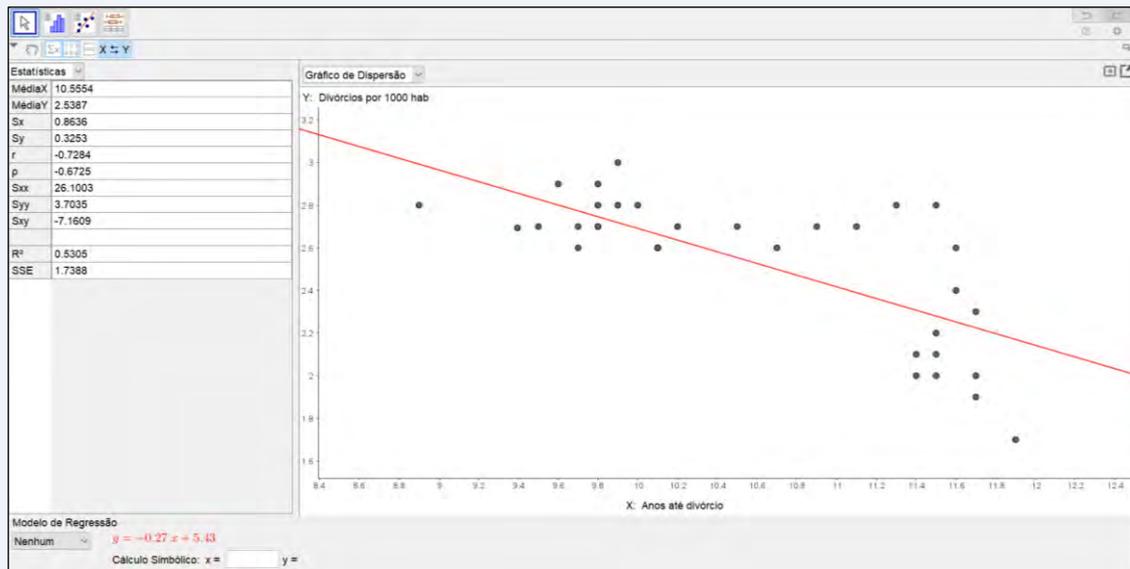
No GeoGebra, foram representados os pontos ($n = 36$) correspondentes aos pares (x, y) , bem como traçadas as retas $x = 10,56$ e $y = 2,54$ que representam a média de cada conjunto de dados, isto é, $\bar{x} = 10,56$ e $\bar{y} = 2,54$.

²⁹ Dados disponíveis em: <https://ourworldindata.org/marriages-and-divorces>



No gráfico, é possível visualizar os pontos marcados a preto que se encontram nos quadrantes em que os produtos $(x - \bar{x})(y - \bar{y})$ são positivos e os pontos a azul que se encontram nos quadrantes em que os produtos $(x - \bar{x})(y - \bar{y})$ são negativos. Observa-se que o número de parcelas negativas é agora maior do que o número de parcelas positivas.

Foi igualmente obtida a equação da reta de regressão $y = -0,27x + 5,43$ e o valor do coeficiente de correlação, $r = -0,7284$, que significa uma correlação linear negativa, moderada/forte entre o número de divórcios por 1000 habitantes e a duração dos casamentos até ao divórcio.

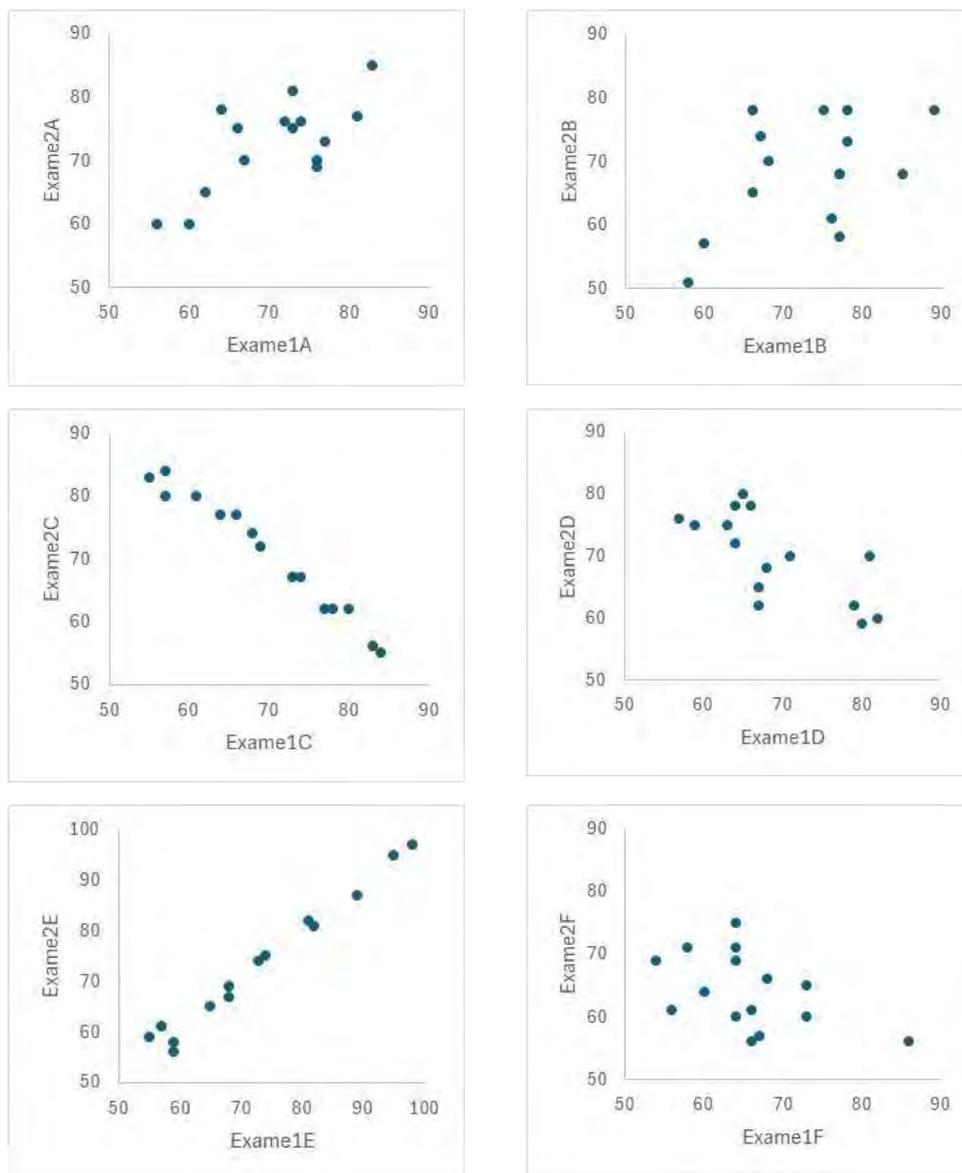


TAREFA³⁰

PARTE I - Interpretação do coeficiente de correlação

Dado um conjunto de dados, o cálculo do coeficiente de correlação, como medida de associação entre duas variáveis, pode causar-nos algumas surpresas, dando-nos informação errada sobre essa associação. Efetivamente, *nem sempre o que parece, é!*

Considera os seguintes diagramas de dispersão correspondentes aos resultados de 2 exames (exame 1 e exame 2) em cada uma de 6 turmas (A a F):



1. Tendo em consideração as representações anteriores, preenche a seguinte tabela, indicando em cada célula, as turmas que verificam a situação referida:

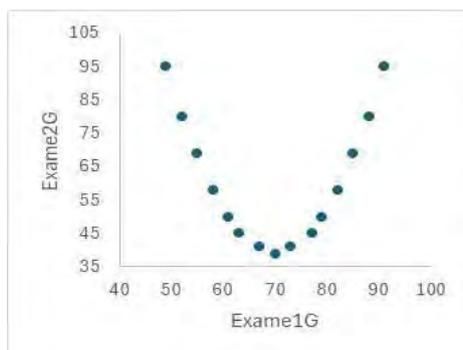
³⁰ Esta tarefa foi adaptada de ActivALEA 4 – Associação ente variáveis quantitativas: o coeficiente de correlação. Disponível em:
https://www.alea.pt/index.php?option=com_content&view=article&id=272&Itemid=1651&lang=pt

Associação	Forte	Moderada	Fraca
Positiva			
Negativa			

Compara a tabela que preencheste com os resultados da tabela seguinte, e verifica se as tuas suposições se confirmam:

Turma	Valor do coeficiente de correlação
A	0.71
B	0.47
C	-0.99
D	-0.72
E	0.99
F	-0.47

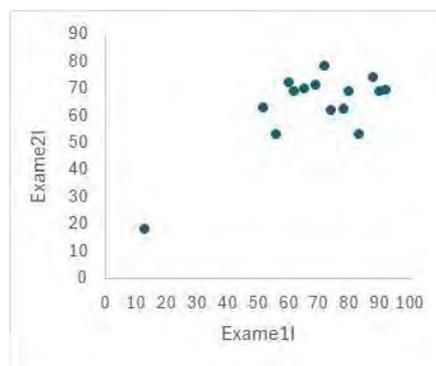
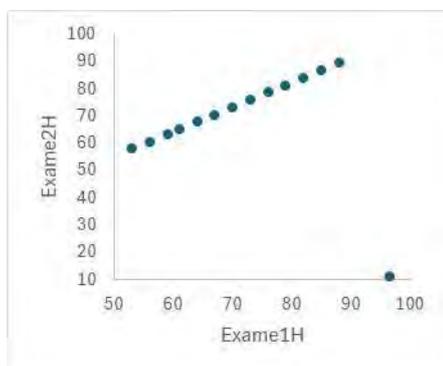
2. Considera agora a seguinte representação correspondente aos dados da turma G:



2.1. Existe algum tipo de associação entre os resultados dos dois exames?

2.2. O valor do coeficiente de correlação linear entre os 2 exames é igual a 0. Como explicas este valor, face à forte associação entre os dois exames da turma G?

3. Considera ainda as representações correspondentes às notas obtidas nas turmas H e I:

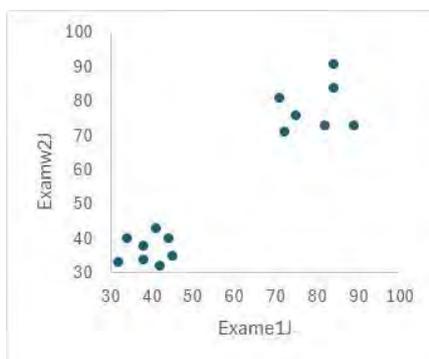


3.1. Para os dois conjuntos de dados relativos às turmas H e I obtiveram-se os coeficientes de correlação 0.70 e 0.04. Qual destes valores é o coeficiente de correlação para a turma H? Justifica.

3.2. Supõe que à turma H se retirava o dado (98, 12) e à turma I se retirava o dado (15, 15). Indica o grau (elevado, moderado, fraco) e o tipo (positivo, negativo) que te parece razoável para os coeficientes de correlação dos dados referentes às turmas H e I, uma vez excluídos os dados que foram retirados. Justifica.

3.3. O facto de na fórmula do coeficiente de correlação entrar a média, faz com que o coeficiente de correlação não seja uma medida resistente. Explica o significado deste termo.

4. Finalmente, considera o seguinte diagrama de dispersão correspondente à turma J:



Se tivesses de escolher entre os valores 0.25, 0.60 e 0.90 para o coeficiente de correlação dos dados da turma J, qual o valor que escolherias? Justifica.

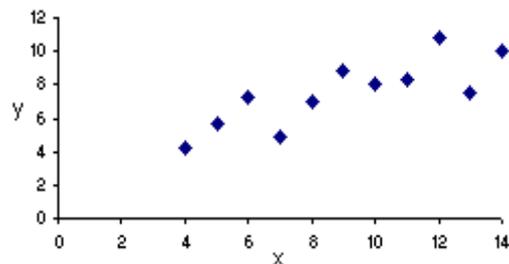
Nota: Antes de calculares e de tentares interpretar o coeficiente de correlação entre duas variáveis, constrói um diagrama de dispersão. Não te esqueças de que o coeficiente de correlação só mede a intensidade com que duas variáveis se associam linearmente, portanto se a representação gráfica não mostrar evidência de associação linear, não tem sentido calculá-lo.

PARTE II – A importância do diagrama de dispersão

Para reforçar o que foi estudado na Parte I, vamos considerar o famoso **quarteto de Anscombe**³¹. Este é um exemplo clássico ([Anscombe](#)) da importância que os gráficos têm no estudo de um conjunto de dados:

1. Considera o seguinte conjunto de dados bivariados e o respetivo diagrama de dispersão:

x	y
10.00	8.04
8.00	6.95
13.00	7.58
9.00	8.81
11.00	8.33
14.00	9.96
6.00	7.24
4.00	4.26
12.00	10.84
7.00	4.82
5.00	5.68



1.1. O que sugere o diagrama de dispersão quanto ao tipo de associação entre as variáveis x e y e quanto à existência de *outliers*?

1.2. Utilizando tecnologia, calcula \bar{x} , \bar{y} e o coeficiente de correlação. O valor obtido para o coeficiente de correlação confirma o tipo de associação que indicaste na alínea anterior?

2. Considera agora os três conjuntos de dados, cujas medidas estatísticas estão a seguir:

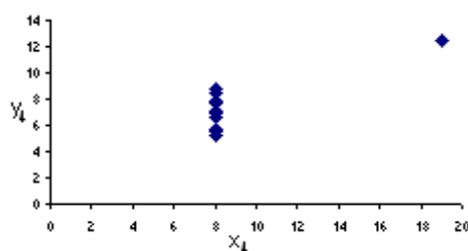
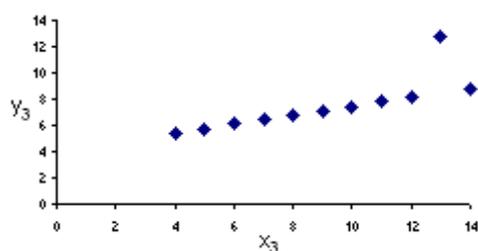
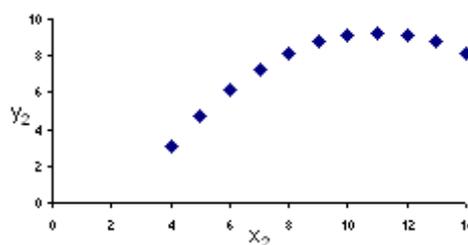
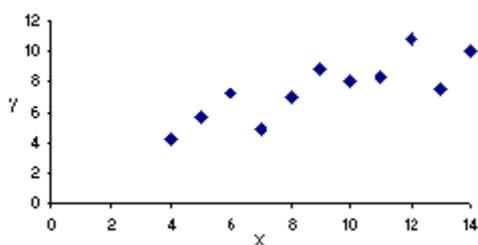
x_2	y_2	x_3	y_3	x_4	y_4
10.00	9.14	10.00	7.46	8.00	6.58
8.00	8.14	8.00	6.77	8.00	5.76
13.00	8.74	13.00	12.74	8.00	7.71
9.00	8.77	9.00	7.11	8.00	8.84
11.00	9.26	11.00	7.81	8.00	8.47
14.00	8.10	14.00	8.84	8.00	7.04
6.00	6.13	6.00	6.08	8.00	5.25
4.00	3.10	4.00	5.39	19.00	12.50
12.00	9.13	12.00	8.15	8.00	5.56
7.00	7.26	7.00	6.42	8.00	7.91
5.00	4.74	5.00	5.73	8.00	6.89

³¹ Adaptado de <http://www.itl.nist.gov/div898/handbook/eda/section1/eda16.htm>

Conjunto de dados 2	Conjunto dados 3	Conjunto dados 4
$n = 11$	$n = 11$	$n = 11$
$\bar{x}_2 = 9.0$	$\bar{x}_3 = 9.0$	$\bar{x}_4 = 9.0$
$\bar{y}_2 = 7.5$	$\bar{y}_3 = 7.5$	$\bar{y}_4 = 7.5$
Correlação $(x_2, y_2) = 0.816$	Correlação $(x_3, y_3) = 0.816$	Correlação $(x_4, y_4) = 0.817$

2.1. Terá sentido dizer que os quatro conjuntos de dados são equivalentes?

2.2. Considera agora os diagramas de dispersão dos quatro conjuntos de dados:



- i) Manténs a resposta que deste na alínea anterior? Justifica.
- ii) Quais as principais conclusões que podes retirar das representações anteriores, quanto aos conjuntos de dados representados?

Nota: Os exemplos anteriores mostram que as estatísticas que utilizamos para reduzir a informação contida nos dados são úteis, mas dão uma visão muito incompleta e limitada dos dados. Elas reduzem drasticamente a informação contida nos dados, através de alguns números. Ao fazerem esta redução dos dados, omitem aspetos importantes e cruciais, pelo que, na melhor situação podemos dizer que dão informação incompleta, mas na pior situação podemos dizer mesmo que dão informação errada.

PARTE III – Interpretação geométrica da correlação linear

Considera agora os dados da tabela seguinte³², que se referem às idades do ator e da atriz que foram vencedores, respetivamente, do óscar de melhor ator principal e do óscar de melhor atriz principal, desde 1929 até 2024. No ano de 1933 houve dois vencedores do óscar de melhor ator principal e no ano de 1966 houve duas vencedoras do óscar de melhor atriz principal. Nesses casos, foi considerada a média das idades dos(as) vencedores(as).

Ano	Idade Melhor Ator	Nome	Ano	Idade Melhor Atriz	Nome
1929	44	Emil Jannings	1929	22	Janet Gaynor
1930	40	Warner Baxter	1930	37	Mary Pickford
1931	62	George Arliss	1931	28	Norma Shearer
1932	53	Lionel Barrymore	1932	63	Marie Dressler
1933	41	Fredric March /Wallace Beery	1933	32	Helen Hayes
1934	34	Charles Laughton	1934	26	Katharine Hepburn
1935	33	Clark Gable	1935	31	Claudette Colbert
1936	52	Victor McLaglen	1936	27	Bette Davis
1937	41	Paul Muni	1937	26	Luise Rainer
1938	37	Spencer Tracy	1938	27	Luise Rainer
1939	38	Spencer Tracy	1939	30	Bette Davis
1940	34	Robert Donat	1940	26	Vivien Leigh
1941	32	James Stewart	1941	29	Ginger Rogers
1942	40	Gary Cooper	1942	24	Joan Fontaine
1943	43	James Cagney	1943	38	Greer Garson
1944	48	Paul Lukas	1944	24	Jennifer Jones
1945	41	Bing Crosby	1945	29	Ingrid Bergman
1946	40	Ray Milland	1946	33	Joan Crawford
1947	49	Fredric March	1947	30	Olivia de Havilland
1948	56	Ronald Colman	1948	34	Loretta Young
1949	41	Laurence Olivier	1949	34	Jane Wyman
1950	38	Broderick Crawford	1950	33	Olivia de Havilland
1951	41	Jose Ferrer	1951	29	Judy Holliday
1952	52	Humphrey Bogart	1952	38	Vivien Leigh
1953	51	Gary Cooper	1953	54	Shirley Booth
1954	35	William Holden	1954	24	Audrey Hepburn
1955	30	Marlon Brando	1955	25	Grace Kelly
1956	38	Ernest Borgnine	1956	47	Anna Magnani
1957	41	Yul Brynner	1957	41	Ingrid Bergman
1958	43	Alec Guinness	1958	27	Joanne Woodward
1959	48	David Niven	1959	41	Susan Hayward
1960	36	Charlton Heston	1960	38	Simone Signoret
1961	47	Burt Lancaster	1961	28	Elisabeth Taylor
1962	31	Maximilian Schell	1962	27	Sophia Loren

³² Dados disponíveis em: <https://www.statcrunch.com/app/index.html?dataid=4513159>

1963	46	Gregory Peck	1963	31	Anne Bancroft
1964	36	Sidney Poitier	1964	37	Patricia Neal
1965	56	Rex Harrison	1965	29	Julie Andrews
1966	41	Lee Marvin	1966	25	Julie Christie
1967	44	Paul Scofield	1967	34	Elizabeth Taylor
1968	42	Rod Steiger	1968	60	Katharine Hepburn
1969	43	Cliff Robertson	1969	44	Barbra Streisand/Katharine Hepburn
1970	62	John Wayne	1970	35	Maggie Smith
1971	43	George C. Scott	1971	34	Glenda Jackson
1972	41	Gene Hackman	1972	34	Jane Fonda
1973	48	Marlon Brando	1973	26	Liza Minnelli
1974	48	Jack Lemmon	1974	37	Glenda Jackson
1975	56	Art Carney	1975	42	Ellen Burstyn
1976	38	Jack Nicholson	1976	41	Louise Fletcher
1977	60	Peter Finch	1977	35	Faye Dunaway
1978	30	Richard Dreyfuss	1978	31	Diane Keaton
1979	40	Jon Voight	1979	41	Jane Fonda
1980	42	Dustin Hoffman	1980	33	Sally Field
1981	37	Robert De Niro	1981	31	Sissy Spacek
1982	76	Henry Fonda	1982	74	Katharine Hepburn
1983	39	Ben Kingsley	1983	33	Meryl Streep
1984	52	Robert Duvall	1984	49	Shirley MacLaine
1985	45	F. Murray Abraham	1985	38	Sally Field
1986	35	William Hurt	1986	61	Geraldine Page
1987	61	Paul Newman	1987	21	Marlee Matlin
1988	43	Michael Douglas	1988	41	Cher
1989	51	Dustin Hoffman	1989	26	Jodie Foster
1990	32	Daniel Day-Lewis	1990	80	Jessica Tandy
1991	42	Jeremy Irons	1991	42	Kathy Bates
1992	54	Anthony Hopkins	1992	29	Jodie Foster
1993	52	Al Pacino	1993	33	Emma Thompson
1994	37	Tom Hanks	1994	35	Holly Hunter
1995	38	Tom Hanks	1995	45	Jessica Lange
1996	31	Nicolas Cage	1996	49	Susan Sarandon
1997	45	Geoffrey Rush	1997	39	Frances McDormand
1998	60	Jack Nicholson	1998	34	Helen Hunt
1999	46	Roberto Benigni	1999	26	Gwyneth Paltrow
2000	40	Kevin Spacey	2000	25	Hilary Swank
2001	36	Russell Crowe	2001	33	Julia Roberts
2002	47	Denzel Washington	2002	35	Halle Berry
2003	29	Adrien Brody	2003	35	Nicole Kidman
2004	43	Sean Penn	2004	28	Charlize Theron
2005	37	Jamie Foxx	2005	30	Hilary Swank
2006	38	Philip Hoffman	2006	29	Reese Witherspoon
2007	45	Forest Whitaker	2007	61	Helen Mirren
2008	50	Daniel Day-Lewis	2008	32	Marion Cotillard
2009	48	Sean Penn	2009	33	Kate Winslet

2010	60	Jeff Bridges	2010	45	Sandra Bullock
2011	50	Colin Firth	2011	29	Natalie Portman
2012	39	Jean Dujardin	2012	62	Meryl Streep
2013	55	Daniel Day-Lewis	2013	22	Jennifer Lawrence
2014	44	Matthew McConaughey	2014	44	Cate Blanchett
2015	32	Eddie Redmayne	2015	54	Julianne Moore
2016	41	Leonardo Di Caprio	2016	26	Brie Larson
2017	41	Casey Affleck	2017	28	Emma Stone
2018	59	Gary Oldman	2018	60	Frances McDormand
2019	37	Rami Malek	2019	44	Olivia Colman
2020	45	Joaquin Phoenix	2020	50	Renée Zellweger
2021	83	Anthony Hopkins	2021	63	Frances McDormand
2022	53	Will Smith	2022	44	Jessica Chastain
2023	54	Brendan Fraser	2023	60	Michelle Yeoh
2024	47	Cillian Murphy	2024	35	Emma Stone

1. Constrói o diagrama de dispersão para as variáveis: *idade do melhor ator* e *idade da melhor atriz*. Com base no diagrama de dispersão, estabelece uma conjectura sobre a associação linear entre as duas variáveis.

2. Determina a média de cada conjunto de dados e representa no gráfico as retas que representam essas médias, de equações $x = \bar{x}$ e $y = \bar{y}$. Observando o modo como os pontos se distribuem pelos quatro quadrantes, o que podes dizer acerca da tua conjectura inicial?

Nota: Se não existe qualquer tipo de associação linear entre as variáveis x e y , então tem-se um valor de r próximo de 0. Neste caso, tanto podem surgir produtos negativos como positivos, distribuindo-se de forma mais ou menos equitativa. Então o valor de r vem próximo de zero.

3. Recorrendo à tecnologia, calcula o coeficiente de correlação linear e interpreta o seu significado na situação considerada.

Bibliografia:

Rossman, A. (1996). *Workshop Statistics, Discovery with data*. Springer.

NOTAS PARA O PROFESSOR

A exploração da tarefa **O coeficiente de correlação linear** é adequada às Aprendizagens Essenciais de Matemática A, Matemática B, MACS e Matemática dos Cursos Profissionais.

A tarefa foca-se na compreensão da importância do coeficiente de correlação linear, como medida de associação entre duas variáveis quantitativas, evidenciando que esta não pode ser a única medida a ser calculada para a análise da forma, direção e grau da associação, uma vez que pode causar-nos algumas surpresas, dando-nos informação errada sobre essa associação. A resolução das três partes da tarefa mobiliza o pensamento crítico dos alunos e incentiva uma interpretação do significado da fórmula utilizada para calcular o coeficiente de correlação linear.

Tema: Estatística

Tópicos e subtópicos:

Dados bivariados

- Dados quantitativos
- Diagrama de dispersão
- Coeficiente de correlação linear
- Reta de regressão
 - variável independente ou explanatória
 - variável dependente ou resposta

Objetivos de aprendizagem:

- Reconhecer que, para estudar a associação entre duas variáveis quantitativas de uma população, se observam essas variáveis sobre cada unidade estatística, obtendo-se uma amostra de pares de dados.
- Reconhecer a importância da representação dos dados no diagrama de dispersão, nuvem de pontos, para interpretar a forma, direção e força da associação (linear) entre as duas variáveis.
- Identificar o coeficiente de correlação linear r , como medida dessa direção e grau de associação (linear), e saber que assume valores pertencentes a $[-1, 1]$, dizendo-se com base nesse valor que a correlação é positiva, negativa ou nula. Recorrer à tecnologia para proceder ao cálculo do coeficiente de correlação linear.
- Compreender que no caso de o diagrama de dispersão mostrar uma forte associação linear entre as variáveis, essa associação pode ser descrita pela reta de regressão ou reta dos mínimos quadrados. Utilizar a tecnologia para determinar uma equação da reta de regressão.
- Compreender que não se pode confundir correlação com relação causa-efeito, pois podem existir variáveis “perturbadoras” que podem provocar uma aparente associação entre as variáveis em estudo.

Duração prevista: Três aulas de 50 minutos.

Materiais e recursos:

- Computador, telemóvel.
- Internet, GeoGebra, folha de cálculo.

Estratégias de implementação:

Após uma primeira introdução à possibilidade de associação entre duas variáveis, no que respeita à forma, direção e força da associação linear, organizar os alunos em pares e propor a resolução das Partes I e II da tarefa. O professor deve circular por entre os alunos, esclarecendo dúvidas e incentivando-os a concluírem a resolução. Recomendamos, que no final deste momento, os alunos corrijam no quadro as duas primeiras partes da tarefa e que se saliente o seguinte:

- Antes de calcular e de tentar interpretar o coeficiente de correlação entre duas variáveis, é importante construir um diagrama de dispersão. Referir que o coeficiente de correlação só mede a intensidade com que duas variáveis se associam linearmente, portanto se a representação gráfica não mostrar evidência de associação linear, não tem sentido calculá-lo.

- Na Parte II, os exemplos mostram que as estatísticas que utilizamos para reduzir a informação contida nos dados são úteis, mas dão uma visão muito incompleta e limitada dos dados. Elas reduzem drasticamente a informação contida nos dados, através de alguns números. Ao fazerem esta redução dos dados, omitem aspetos importantes e cruciais, pelo que, na melhor situação podemos dizer que dão informação incompleta, mas na pior situação podemos dizer mesmo que dão informação errada.

Antes de resolver a tarefa é importante que os alunos se apropriem do conteúdo que consta na secção *Para ler e compreender*. Recomendamos que o professor explore com os alunos esta informação.

Posteriormente, mantendo os pares anteriores, os alunos deverão resolver a Parte III. Deverá ser dado tempo aos alunos para que estes possam formular e testar as próprias conjeturas. O papel do professor deverá mediar esta construção das conjeturas, dando um feedback constante. No final da aula, esta parte da tarefa também deverá ser corrigida no quadro pelos alunos, promovendo uma discussão coletiva e deverá ser feita uma sistematização das aprendizagens, reforçando que não se pode confundir correlação com relação causa-efeito, pois podem existir variáveis “perturbadoras” que podem provocar uma aparente associação entre as variáveis em estudo.

Dificuldades esperadas e ações do professor:

Os alunos podem revelar algumas dificuldades, nomeadamente na apropriação de conceitos estatísticos novos, como o coeficiente de correlação ou a importância da construção do diagrama de dispersão para analisar a associação de variáveis. Poderão surgir dificuldades com o uso do GeoGebra, que merecerão o apoio do professor.

Na Parte III, o professor terá um papel importante na promoção da discussão para que os alunos sejam capazes de interpretar geometricamente a fórmula do cálculo do coeficiente de correlação.

O professor deverá ter um papel mediador, prevendo as questões a colocar aos alunos por forma a conduzi-los à formulação da conjectura de que não existe associação entre as duas variáveis em estudo.

Avaliação:

A avaliação deverá ter um carácter formativo, baseando-se no feedback oral durante a resolução da tarefa.

Sugestão:

Um site interessante para observar, de forma dinâmica, a influência de um único ponto na representação da reta de regressão e no coeficiente de correlação está disponível em <https://rpsychologist.com/correlation/>.

PROPOSTA DE RESOLUÇÃO**PARTE I**

1. Analisando os gráficos, a tabela fica assim preenchida:

Associação	Forte	Moderada	Fraca
Positiva	E	A	B
Negativa	C	D	F

Assim, tendo em conta o coeficiente de correlação obtido para cada turma, tem-se:

Turma	Valor do coeficiente de Correlação	Sentido e força
A	0.71	Posit. Mod.
B	0.47	Posit. Fraca
C	-0.99	Negat. Forte
D	-0.72	Negat. Mod.
E	0.99	Posit. Forte
F	-0.47	Negat. Fraca

2.

2.1. A forma da nuvem de pontos, para o caso da turma G, revela que existe uma associação forte entre as duas variáveis, embora não se trate de uma associação linear, mas sim de uma relação que será provavelmente quadrática.

2.2. O facto de o coeficiente de correlação linear ser igual a 0 significa que não existe uma associação linear entre as duas variáveis. No entanto, pode haver outro tipo de correlação forte entre as duas variáveis, nomeadamente quando os pontos se dispõem ao longo do gráfico de uma outra função, como a função quadrática.

3.

3.1. No caso da turma H, todos os pontos, exceto um, estão perfeitamente dispostos sobre uma reta, sugerindo uma correlação linear positiva forte entre as variáveis. É de supor que o único ponto que se afasta da reta seja um *outlier*. No caso da turma I, os pontos não estão claramente alinhados em torno de uma reta, formando uma nuvem dispersa, o que sugere uma correlação muito fraca, existindo ainda um ponto afastado dos restantes. Nestes dois casos o efeito do *outlier* faz com que o coeficiente de correlação da turma H seja 0.04 e o da turma I seja 0.70.

3.2. No caso da turma H, é retirado o ponto que mais se afasta da reta, pelo que todos os outros pontos ficarão alinhados sobre essa reta. Nesse caso, a correlação passará a ser elevada e positiva. No caso da turma I, será excluído um ponto que está mais afastado dos restantes. Assim, os restantes pontos formam uma nuvem dispersa, sugerindo uma correlação fraca e positiva.

3.3. Como sabemos, a média não é uma medida resistente, ou seja, é muito influenciada por valores muito grandes ou muito pequenos relativamente aos restantes, os *outliers*. Dado que o cálculo do coeficiente de correlação envolve a média dos x_i e a média dos y_i , percebe-se que o coeficiente de correlação seja igualmente sensível à existência de *outliers*.

4. O valor 0.95 é aquele que melhor parece descrever a força da correlação positiva entre as variáveis. Este valor indica uma correlação moderada. No gráfico é possível perceber que os pontos estão aproximadamente dispostos ao longo de uma reta, embora formando dois aglomerados.

PARTE II

1.

1.1. O diagrama de dispersão sugere que existe uma correlação linear positiva forte entre as duas variáveis. Não parecem existir *outliers*.

1.2. Usando a folha de cálculo, obtiveram-se os resultados:

$$\bar{x} = 9.00, \quad \bar{y} = 7.50, \quad r = 0.816$$

2.

2.1. Embora as medidas estatísticas, \bar{x} , \bar{y} , r , sejam iguais nos quatro casos, não se pode concluir que os conjuntos de dados são equivalentes, pois nada se sabe sobre a forma como os dados bivariados se associam.

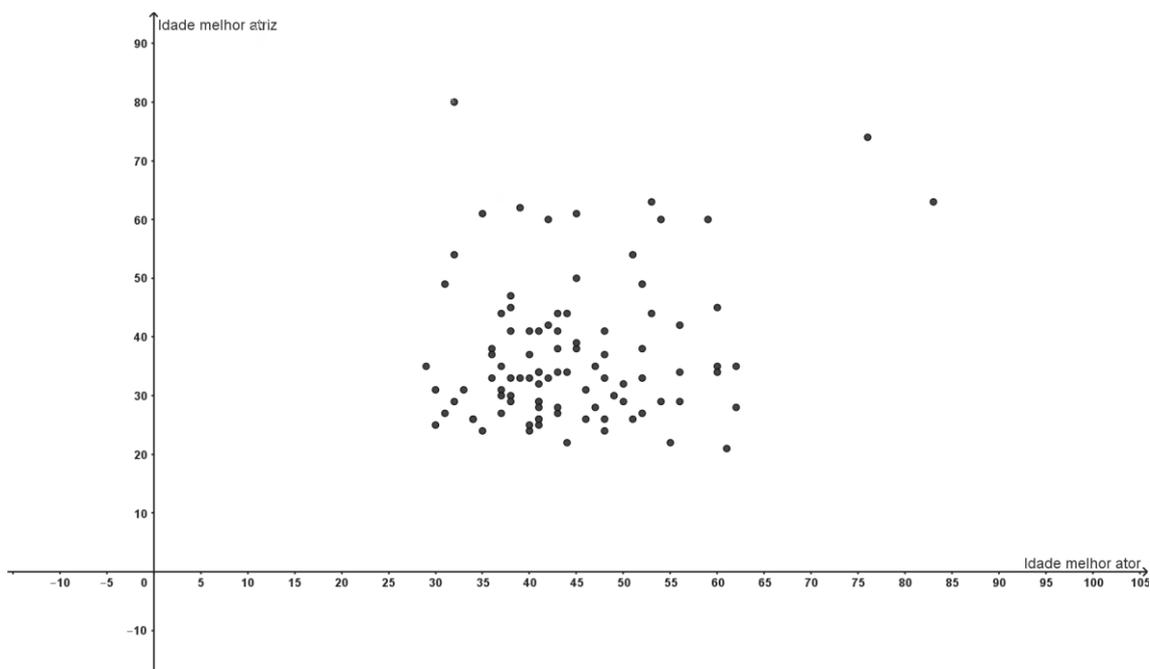
2.2. Considerando os diagramas de dispersão, percebe-se que o tipo de associação entre as variáveis é diferente de caso para caso.

i) Analisando as nuvens de pontos, é claro que os dados não são equivalentes no que se refere ao tipo de associação entre as variáveis.

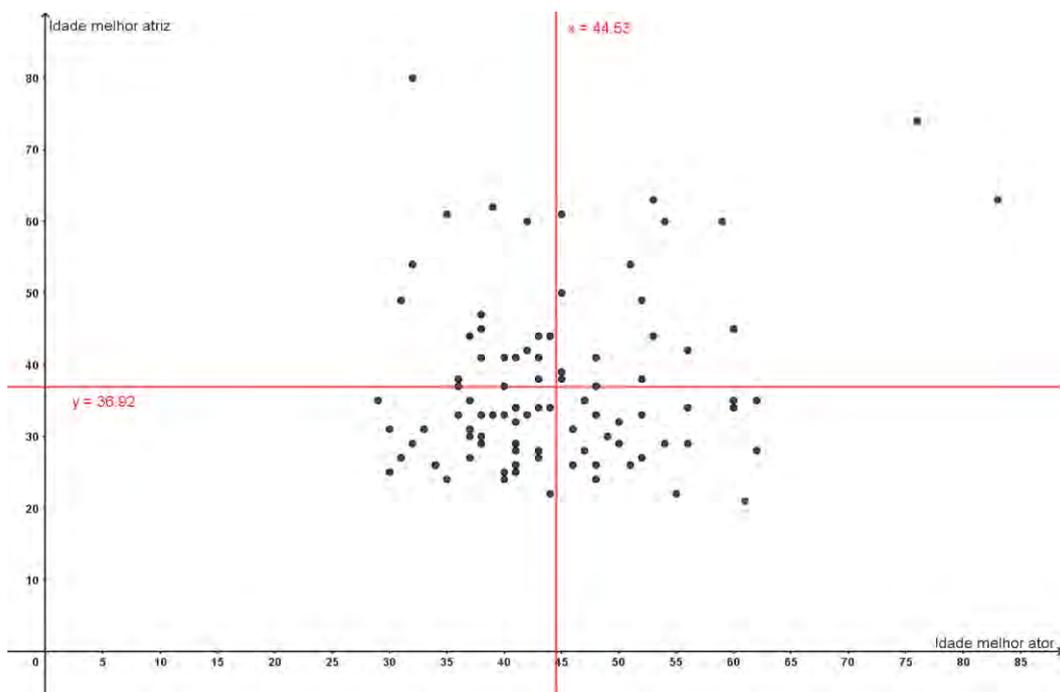
ii) No conjunto de dados 1, observa-se uma associação linear positiva forte; no conjunto de dados 2, não há uma associação linear entre as variáveis, embora exista uma clara associação entre as variáveis (provavelmente quadrática); no conjunto de dados 3, há uma associação linear positiva forte e aparentemente existe um *outlier*, sem o qual a correlação seria muito forte; no conjunto de dados 4 não há associação linear entre as variáveis.

PARTE III

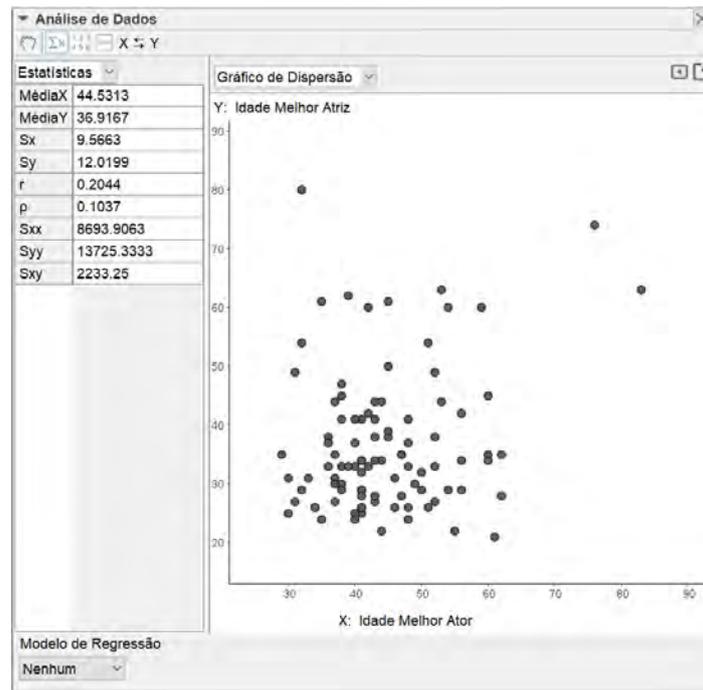
1. O diagrama de dispersão (obtido com o GeoGebra) indica que não existe uma associação linear entre as duas variáveis, ou seja, o coeficiente de correlação linear deverá ser próximo de zero.



2. Usando o GeoGebra, foi obtida a média de cada conjunto de dados (em anos): $\bar{x} = 44,5$ e $\bar{y} = 36,9$. Foi feita a representação das duas retas que delimitam os quatro quadrantes. Os pontos parecem distribuir-se de forma aproximadamente igual entre os quatro quadrantes, o que reforça a ideia de que a correlação linear será muito fraca.



3. Recorrendo ao GeoGebra, obteve-se o coeficiente de correlação $r = 0,2044$.



É um valor baixo, que indica uma correlação fraca. Isto significa que não se pode falar de uma associação linear entre a idade do vencedor do óscar de melhor ator e a idade da vencedora do óscar de melhor atriz. Esta conclusão parece razoável pois é natural que a idade do ator vencedor não tenha qualquer relação com a idade da atriz vencedora, em cada atribuição dos óscares.

QUANTO MAIS POLUIÇÃO, MELHOR?

TAREFA



Uma associação forte entre as duas variáveis, de um par de variáveis, não implica uma relação de causa-efeito entre as variáveis.

Uma das ideias mais importantes a ter presente, quando se estuda a associação entre as variáveis X e Y do par (X,Y) , é o facto de **uma associação forte entre as variáveis não implicar uma relação de causa-efeito entre elas**. Vamos a seguir explorar esta situação.

A tabela seguinte apresenta, para alguns países, a informação sobre a emissão anual de CO_2 (em toneladas) *per capita* e a esperança de vida (em anos), em 2021 (dados recolhidos do site <https://ourworldindata.org/>):

País	Emissão, <i>per capita</i> , de CO_2	Esperança vida	País	Emissão, <i>per capita</i> , de CO_2	Esperança vida
Afghanistan	0,3	62	Finland	6,8	82
Algeria	4	76,4	France	4,7	82,5
Andorra	5,7	80,4	Gabon	2,4	65,8
Angola	0,6	61,6	Gambia	0,2	62,1
Argentina	4,1	75,4	Georgia	2,9	71,7
Armenia	2,4	72	Germany	8,1	80,6
Austria	7,2	81,6	Ghana	0,6	63,8
Azerbaijan	3,7	69,4	Greece	5,4	80,1
Bahamas	5,9	71,6	Grenada	2,6	74,9
Belgium	8,2	81,9	Guatemala	1,2	69,2
Benin	0,6	59,8	Guinea	0,4	58,9
Bolivia	1,9	63,6	Guinea-Bissau	0,2	59,7

Bosnia and Herzegovina	4,1	75,3	Haiti	0,3	63,2
Botswana	2,5	61,1	Hungary	5	74,5
Brazil	2,3	72,8	Iceland	9,1	83,7
Burkina Faso	0,3	59,3	India	1,9	67,2
Burundi	0,1	61,7	Ireland	7,5	82
Cambodia	1,1	69,6	Israel	6,1	82,3
Cameroon	0,3	60,3	Italy	5,5	82,9
Central African Republic	0	53,9	Japan	8,6	84,8
Chile	4,4	78,9	Kenya	0,4	61,4
Colombia	1,8	72,8	Liberia	0,2	60,7
Comoros	0,4	63,4	Mexico	3,2	70,2
Congo	1,3	63,5	Mozambique	0,2	59,3
Cote d'Ivoire	0,4	58,6	Netherlands	8,1	81,7
Croatia	4,4	77,6	New Zealand	6,6	82,5
Cuba	2	73,7	Norway	7,6	83,2
Cyprus	6,1	81,2	Peru	1,7	72,4
Dem. Republic of Congo	0	59,2	Philippines	1,3	69,3
Djibouti	0,3	62,3	Portugal	4	81
Dominica	2,2	72,8	Sao Tome and Principe	0,6	67,6
Dominican Republic	2,6	72,6	Singapore	5,5	82,8
East Timor	0,7	67,7	Slovakia	6,5	74,9
Ecuador	2,3	73,7	Slovenia	5,9	80,7
Egypt	2,3	70,2	Spain	4,9	83
El Salvador	1,1	70,7	United Kingdom	5,2	80,7
Ethiopia	0,1	65	Zimbabwe	0,7	59,3
Fiji	1,6	67,1			

1. Na tabela dada, identifica:

1.1. Os 2 países mais poluidores e respetivo continente; os 2 países com maior esperança de vida e respetivo continente;

1.2. Os 2 países menos poluidores e respetivo continente; os 2 países com menor esperança de vida e respetivo continente.

1.3. Notas algum padrão nos valores registados? Justifica a tua resposta.

2. Representa num diagrama de dispersão o par de variáveis (*Emissão de CO₂*, *Esperança de vida*). Qual o aspeto da nuvem de pontos?

3. Se adequado, calcula o coeficiente de correlação linear entre as variáveis (*Emissão de CO₂*, *Esperança de vida*).

4. Tendo em consideração o valor do coeficiente de correlação obtido na alínea anterior, classifica o tipo de associação entre as variáveis *Emissão de CO₂* e *Esperança de vida*, quanto ao grau e direção.

5. Obtém o modelo de regressão linear (reta dos mínimos quadrados) que se ajusta aos dados.

6. Explora a informação sobre a emissão de CO₂ relativa a outros países não considerados na tabela anterior (<https://ourworldindata.org/co2-emissions>) e diz se poderias estimar, a partir da reta de regressão, a esperança de vida dos Estados Unidos da América, em 2021 (verifica que a emissão de CO₂ é 16,5). Justifica.

7. Pensas que o grau da associação que obtiveste entre as variáveis, assim como o modelo de regressão linear, te permite estabelecer uma relação de causa-efeito entre as variáveis em estudo? Será que tendo em conta o grau de associação, faz sentido admitir que fazendo mais poluição, aumenta a esperança de vida? Justifica, devidamente, a tua resposta.

NOTAS PARA O PROFESSOR

A exploração da tarefa **Quanto mais poluição melhor?** é adequada às Aprendizagens Essenciais de Matemática A, Matemática B, MACS e Matemática dos Cursos Profissionais.

A tarefa foca-se na compreensão de que uma associação forte entre as duas variáveis de um par de variáveis não implica uma relação de causa-efeito entre as variáveis. De facto, nesta tarefa estamos perante uma situação em que não se pode estabelecer uma relação de causa-efeito entre as variáveis *Emissões de CO₂* e *Esperança de vida*, pois embora elas cresçam aparentemente em associação, existe uma terceira variável, a que podemos chamar o *Nível de industrialização do país*, associado à qualidade de vida, que faz aumentar as emissões de CO₂ e a esperança de vida!

Verifica-se aqui uma correlação a que se costuma dar o nome de correlação espúria. O que é importante ter presente é que correlação não significa causalidade. Esta tarefa contribui para o desenvolvimento do pensamento crítico dos alunos.

Tema: Estatística

Tópicos e subtópicos:

Dados bivariados

- Dados quantitativos
- Diagrama de dispersão
- Coeficiente de correlação linear
- Reta de regressão
 - variável independente ou explanatória
 - variável dependente ou resposta

Objetivos de aprendizagem:

- Reconhecer que, para estudar a associação entre duas variáveis quantitativas de uma população, se observam essas variáveis sobre cada unidade estatística, obtendo-se uma amostra de pares de dados.
- Reconhecer a importância da representação dos dados no diagrama de dispersão, nuvem de pontos, para interpretar a forma, direção e força da associação (linear) entre as duas variáveis.
- Identificar o coeficiente de correlação linear r , como medida dessa direção e grau de associação (linear), e saber que assume valores pertencentes a $[-1, 1]$, dizendo-se com base nesse valor que a correlação é positiva, negativa ou nula. Recorrer à tecnologia para proceder ao cálculo do coeficiente de correlação linear.
- Compreender que no caso de o diagrama de dispersão mostrar uma forte associação linear entre as variáveis, essa associação pode ser descrita pela reta de regressão ou reta dos mínimos quadrados. Utilizar a tecnologia para determinar uma equação da reta de regressão.

- Compreender que não se pode confundir correlação com relação causa-efeito, pois podem existir variáveis “perturbadoras” que podem provocar uma aparente associação entre as variáveis em estudo.

Duração prevista: Duas aulas de 50 minutos.

Materiais e recursos:

- Computador, telemóvel.
- Internet, GeoGebra, Folha de cálculo.

Estratégias de implementação:

Sugere-se que o professor organize os alunos em pares e proponha a resolução da tarefa. Não é necessária qualquer introdução teórica, porque na data da aplicação desta tarefa os alunos já estarão familiarizados com a análise de pares de variáveis recorrendo ao diagrama de dispersão e ao coeficiente de correlação.

No final da resolução da tarefa, por parte dos alunos, o professor deve proceder à sua correção no quadro, e promover a discussão em grande grupo, deixando claro que não se pode estabelecer uma relação de causa-efeito entre as variáveis *Emissões de CO₂* e *Esperança de vida*, pois embora elas cresçam aparentemente em associação, existe uma terceira variável, a que podemos chamar o *Nível de industrialização do país*, associado à qualidade de vida, que faz aumentar as emissões de CO₂ e a esperança de vida! A este tipo de correlação, costuma chamar-se correlação espúria.

É importante destacar que correlação não significa causalidade.

Dificuldades esperadas e ações do professor:

Os alunos deverão ter muita autonomia na realização desta tarefa e já se deverão ter apropriado dos conhecimentos estatísticos e do uso da tecnologia necessária, para a sua realização. O professor deverá ter um papel mediador, disponibilizando feedback oral e conduzindo os alunos na realização e conclusão da tarefa.

Avaliação:

A avaliação deverá ter um carácter formativo, recorrendo ao feedback oral durante a resolução da tarefa.

PROPOSTA DE RESOLUÇÃO

Para responder às questões relativas aos dados apresentados, o primeiro passo será copiar a tabela anterior para uma folha de cálculo. Nesta resolução utilizou-se o Excel:

	A	B	C
1	País	Emissão, per capita, de CO ₂ em 2021	Esperança vida em 2021
2	Afghanistan	0,3	62
3	Algeria	4	76,4
4	Andorra	5,7	80,4
5	Angola	0,6	61,6
6	Argentina	4,1	75,4
7	Armenia	2,4	72
8	Austria	7,2	81,6
9	Azerbaijan	3,7	69,4
10	Bahamas	5,9	71,6
11	Belgium	8,2	81,9
12

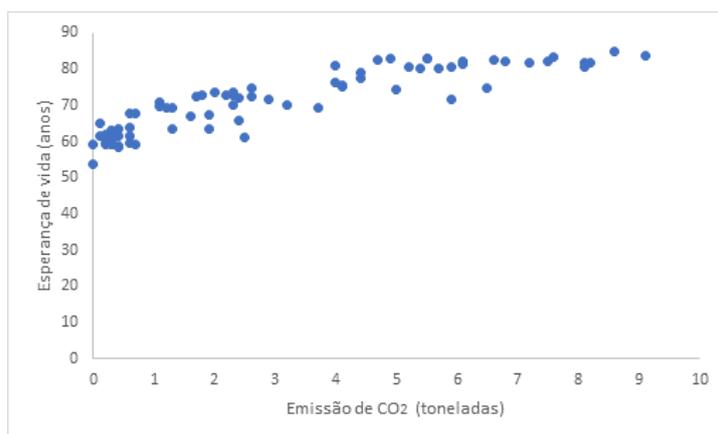
1. Ordenando a tabela, utilizando como critério a coluna da variável *Emissão de CO₂*, e de seguida a coluna da variável *Esperança de vida*, obtêm-se as respostas às questões da alínea a), nomeadamente:

1.1. Países mais poluidores: **Japão** no continente asiático e **Islândia** no continente europeu; Países com maior esperança de vida: **Japão** no continente asiático e **Islândia** no continente europeu;

1.2. Países menos poluidores: **República Centro-Africana** e **República Democrática do Congo** no continente africano; Países com menor esperança de vida: **República Centro-Africana** e **Costa do Marfim**, no continente africano.

1.3. Verifica-se que, de um modo geral, os países mais poluidores têm maior esperança de vida e os menos poluidores têm menor esperança de vida. Embora a República Democrática do Congo não seja um dos dois com menor esperança de vida, é um dos quatro com menor esperança de vida, de entre os países considerados no estudo.

2. O diagrama de dispersão obtém-se diretamente a partir do Excel e apresenta o seguinte aspeto:



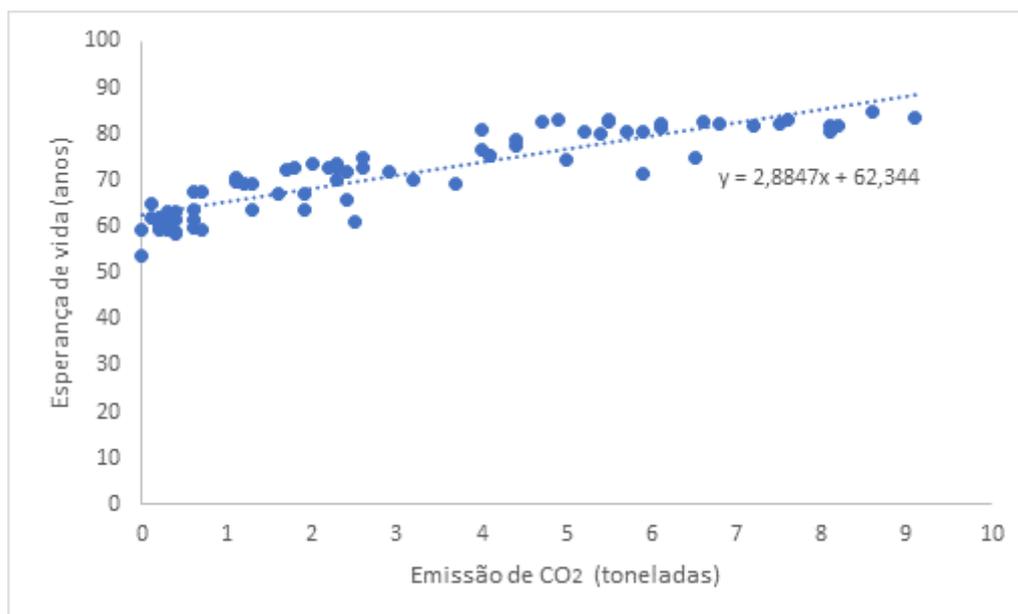
A nuvem de pontos mostra um aspeto alongado, aparentando uma associação linear positiva entre as variáveis consideradas.

3. Tendo em consideração que o diagrama de dispersão apresenta uma associação linear entre as variáveis, calcula-se o coeficiente de correlação linear, já que é esta a medida que mede a força e

direção com que as variáveis se associam. O valor obtido para o coeficiente de correlação linear, representado por r , é $r \approx 0,90$.

4. O valor do coeficiente de correlação linear obtido na alínea anterior indica que existe uma forte associação linear positiva entre as variáveis em estudo, isto é, à medida que as emissões de CO_2 sobem, de um modo geral, também sobe a esperança de vida.

5. O modelo linear é dado pela reta, cuja equação se apresenta no gráfico seguinte, onde x representa a variável *Emissão de CO_2* e y representa a variável *Esperança de vida*:



6. A partir da página indicada, verifica-se que a emissão de CO_2 , *per capita*, nos Estados Unidos da América (EUA) em 2021 foi de 16,5 toneladas, valor este que está fora do intervalo das emissões consideradas, que é $[0;9,1]$, pelo que não se deve utilizar a reta de regressão para estimar o valor da esperança de vida nos EUA, em 2021. A título de curiosidade, se utilizássemos a reta de regressão linear para estimar a esperança de vida nos EUA em 2021, obteríamos um valor aproximadamente igual a 110 anos ($= 2,8847 \times 16,5 + 62,344$), que neste momento, ainda está fora da realidade.

7. Estamos perante uma situação em que não se pode estabelecer uma relação de causa-efeito entre as variáveis *Emissões de CO_2* e *Esperança de vida*, pois embora elas cresçam aparentemente em associação, existe uma terceira variável, a que podemos chamar o *Nível de industrialização* do país, associado à qualidade de vida, que faz aumentar as *Emissões de CO_2* e a *Esperança de vida*! A uma variável deste tipo, é usual chamar *variável perturbadora (confounding variable)*, pois provoca uma forte correlação entre duas variáveis, que não tem qualquer significado. Verifica-se aqui uma correlação a que se costuma dar o nome de *correlação espúria*. O que é importante ter presente é que correlação não significa causalidade.

Nota: Nesta situação foi fácil identificar uma *variável perturbadora*, que provoca uma *correlação espúria* entre duas variáveis. No entanto, uma *correlação espúria* pode-se verificar simplesmente devido ao acaso, sem a intervenção de uma *variável perturbadora*.

NA HORA DE SUBSTITUIR OS PNEUS DO CARRO

TAREFA



PARTE I

O Automóvel Clube de Portugal (ACP) alerta para a importância de substituir os pneus dos carros, explicando³³:

“Os cuidados a ter com os pneus permitem prolongar-lhes a vida útil. Porém, todos os pneus acabam por sofrer desgaste, sendo necessário substituí-los. Regra geral, devem ser trocados de 5 em 5 anos (ou, no máximo, 10 anos). (...) Na hora de substituir os seus pneus, não olhe apenas ao preço. Lembre-se que a segurança é uma prioridade”.

No site *Ekonomista*, foi publicada em julho de 2020 uma notícia³⁴ sobre a durabilidade dos pneus de diversas marcas disponíveis no mercado:

Um estudo realizado pela Deco Proteste concluiu que a escolha dos pneus corretos pode ajudar o seu carro a percorrer mais 12 mil quilómetros do que a marca com pior avaliação. No total, foram entrevistados quase cinco mil portugueses para recolher a opinião e experiência sobre 25 marcas. Porém, antes de elaborar a análise final, a Deco partilhou dados com as suas congéneres espanholas, italianas e belgas. Após confrontados todos os resultados foram obtidas mais de 37 mil respostas.

A notícia apresenta a conclusão do estudo, revelando uma lista de 25 marcas de pneus e a respetiva durabilidade, em quilómetros feitos, como mostra a tabela

³³ Citado de: <https://www.acp.pt/veiculos/conductor-em-dia/conduzir-em-seguranca/cuidados-a-ter-com-os-pneus>

³⁴ Disponível em: <https://www.e-konomista.pt/pneus-que-duram-mais-quilometros/>

seguinte (**Nota:** na tabela incluímos apenas 24 marcas porque uma delas não foi encontrada no mercado, tendo sido removida nesta proposta de trabalho):

Marca de pneus	Durabilidade (km)
Barum	37504
BF Goodrich	44096
Bridgestone	42944
Continental	43258
Dunlop	42551
Falken	34494
Firestone	40059
Fulda	41095
Goodyear	42927
GT Radial	36597
Hankook	41599
Khumo	39149
Kleber	37589
Kormoran	34975
Mabor	35118
Matador	41190
Michelin	46393
Nankang	34210
Pirelli	41021
Semperit	37444
Toyo	41794
Uniroyal	41413
Vredestein	40573
Yokohama	42311

1. Como designarias o tipo de estudo estatístico feito pela DECO, que é referido na notícia? Qual foi a população em estudo?
2. O que podes dizer sobre a representatividade da amostra de inquiridos? Explica a tua resposta.
3. Considerando o objetivo do estudo, classifica a variável estatística analisada.
4. Analisa a distribuição dos dados relativos à durabilidade dos pneus, calculando a média e a mediana.
 - 4.1. Comparando as duas medidas que calculaste anteriormente, o que esperas quanto à forma como os dados se distribuem?

4.2. Uma durabilidade de 40000 km significa, segundo alguns peritos, uma expectativa de vida muito longa dos pneus. É verdade que esse critério é cumprido por, pelo menos, 50% das marcas mencionadas? Justifica.

5. Analisa a distribuição dos dados relativos à durabilidade dos pneus, calculando a amplitude, os quartis e a amplitude interquartil.

5.1. A comparação destas medidas dá-te alguma informação sobre a forma como os dados se distribuem?

6. Para complementares o estudo da distribuição dos dados relativos à durabilidade dos pneus, utiliza uma folha de cálculo para obteres um diagrama em caixa de bigodes.

6.1. Interpreta a representação obtida, não deixando de referir os resultados obtidos nos pontos 4. e 5.

6.2. Qual o aspeto da forma da distribuição dos dados? Simétrica, aproximadamente simétrica, enviesada à direita, à esquerda, etc.?

7. Calcula o desvio padrão amostral e comenta a afirmação seguinte (Sugestão: usa o site ViaMichelin):

Relativamente à durabilidade média dos pneus considerados, há uma dispersão, em quilómetros, aproximadamente igual a uma viagem de Lisboa a Varsóvia (ver mapa).



8. Perante os resultados da análise estatística realizada, qual seria a tua recomendação sobre a marca a escolher para uma troca de pneus? Justifica a tua recomendação.

PARTE II

Como vimos antes, o ACP aconselha: “Na hora de substituir os seus pneus, não olhe apenas ao preço. Lembre-se que a segurança é uma prioridade.”

No estudo feito na Parte I, a variável preço não foi considerada. Porém, é possível fazer uma pesquisa online de preços de pneus de várias marcas.

Foi feita uma consulta (no site <https://www.qualpneu.pt/>) e registado o preço mais baixo encontrado, para cada marca, de 1 pneu do tipo 205/55 R16 91V. Os dados são apresentados na tabela seguinte:

Marca	Preço (€)
Barum	80,69
BF Goodrich	80,80
Bridgestone	77,88
Continental	79,60
Dunlop	72,25
Falken	64,80
Firestone	63,30
Fulda	69,59
Goodyear	75,61
GT Radial	72,58
Hankook	74,51
Khumo	56,70
Kleber	72,83
Kormoran	52,15
Mabor	71,06
Matador	60,70
Michelin	104,90
Nankang	46,70
Pirelli	76,60
Semperit	75,68
Toyo	74,53
Uniroyal	64,60
Vredestein	66,46
Yokohama	63,20

1. Faz uma análise da distribuição dos preços dos pneus das várias marcas, incluindo na tua análise medidas de localização e de dispersão que poderás obter, recorrendo a uma calculadora gráfica ou a uma folha de cálculo.
2. Constrói na folha de cálculo um diagrama em caixa de bigodes e um histograma dos dados relativos aos preços dos pneus.

2.1. Interpreta as duas representações gráficas obtidas no contexto da distribuição dos dados. Encontre algum aspeto dessa distribuição que te mereça referência especial, nomeadamente a existência de algum *outlier*?

2.2. Se encontraste *outliers*, retira-os do conjunto dos dados e calcula as seguintes diferenças: (mediana – mínimo) e (máximo – mediana). O que podes concluir sobre a relação entre as amplitudes dos 50% dos preços mais baixos e dos 50% dos preços mais altos? Como é traduzida nas representações gráficas esta diferença de amplitudes?

3. Averigua se existe ou não uma associação linear entre a durabilidade e o preço dos pneus, considerando a durabilidade como variável explanatória (ou independente) e o preço como a variável resposta (ou dependente). Para isso, constrói um diagrama de dispersão e analisa a nuvem de pontos. Se adequado, ajusta uma reta de regressão e calcula o coeficiente de correlação linear. Notas a existência de alguns pontos que podem ser *outliers*? O que concluis?

4. Perante os resultados da análise estatística realizada, tens fundamentos para recomendar alguma marca a escolher para uma troca de pneus? Explica a tua resposta.

NOTAS PARA O PROFESSOR

A exploração da tarefa – **Na hora de substituir os pneus do carro** – é adequada às Aprendizagens Essenciais de Matemática A, Matemática B, MACS e Matemática dos Cursos Profissionais.

A tarefa tem um grande foco no desenvolvimento sustentável, bem como na tomada de decisões informadas, na ótica do cidadão e consumidor, enquadrando-se, portanto, numa perspetiva de Matemática para Todos, promovendo a comunicação matemática e o desenvolvimento do espírito crítico.

Tema: Estatística

Tópicos e subtópicos:

População, amostra e variável

- Fases de um procedimento estatístico

Organização de dados

- Histograma
- Medidas de localização
- Medidas de dispersão
- Propriedades das medidas

Dados bivariados

- Diagrama de dispersão
- Coeficiente de correlação linear
- Reta de regressão

Objetivos de aprendizagem:

- Identificar num estudo estatístico, população, amostra e a(s) característica(s) a estudar, que se designa(m) por variável(variáveis).
- Organizar e representar a informação contida em dados quantitativos discretos e contínuos em tabelas e interpretá-las.
- Construir histogramas, considerando classes com a mesma amplitude.
- Selecionar representações gráficas adequadas para cada tipo de dados identificando vantagens/inconvenientes, relembrando a construção de diagramas de extremos e quartis.
- Interpretar as medidas de localização, em particular a média (\bar{x}) e a mediana (Me), na caracterização da distribuição dos dados, relacionando-a com as representações gráficas obtidas.
- Interpretar as medidas de dispersão, amplitude, amplitude interquartil e desvio padrão amostral, s , na caracterização da distribuição dos dados.

- Reconhecer que existem situações em que é preferível utilizar como medida de dispersão a amplitude interquartil em vez do desvio padrão.
- Reconhecer a importância da representação dos dados no diagrama de dispersão ou nuvem de pontos, para interpretar a forma, direção e força da associação (linear) entre as duas variáveis.
- Identificar o coeficiente de correlação linear r , como medida dessa direção e grau de associação (linear), e saber que assume valores pertencentes ao intervalo $[-1,1]$, dizendo-se, com base nesse valor, que a correlação é positiva, negativa ou nula. Recorrer à tecnologia para proceder ao cálculo do coeficiente de correlação linear.
- Compreender que a existência de *outliers* influencia estes procedimentos.

Duração prevista: Três aulas de 50 minutos.

Materiais e recursos:

- Computador, calculadora gráfica, telemóvel.
- Internet, folha de cálculo.

Estratégias de implementação:

O alerta do Automóvel Clube de Portugal sobre a importância de mudar os pneus de um automóvel, por questões de segurança, serve de mote a esta tarefa.

Recomendamos que o professor organize os alunos em pares para a resolução da tarefa. Na primeira parte, recorrendo a conceitos estatísticos, em particular, às medidas de localização e de dispersão, pede-se aos alunos que façam uma análise sobre a durabilidade dos pneus das diferentes marcas.

A segunda parte da tarefa mobiliza os alunos para uma análise do preço dos pneus. Nesta parte da tarefa sugere-se aos alunos que analisem a relação entre a durabilidade dos pneus e o respetivo preço. O facto de a segunda parte da tarefa envolver dados bivariados promove uma reflexão sobre a relação entre as duas variáveis. É importante levar os alunos a reconhecerem a existência de um *outlier* e interpretarem este dado no contexto da aquisição de pneus.

Na fase final deve ser realizada uma síntese das aprendizagens realizadas e das principais conclusões da tarefa.

Dificuldades esperadas e ações do professor:

Os alunos podem revelar algumas dificuldades, nomeadamente ao nível da organização e da interpretação dos dados no contexto da tarefa.

Na Parte II da tarefa podem surgir dificuldades associadas à existência de um valor atípico (*outlier*) evidenciado na representação gráfica.

O professor, sempre que necessário, deve fazer referência a conceitos estatísticos de anos anteriores e ajudar os alunos na utilização da tecnologia (folha de cálculo e/ou calculadora gráfica), dando sugestões ou colocando questões que contribuam para o desenvolvimento do trabalho.

Avaliação:

A avaliação poderá ter um carácter formativo.

Sugere-se que cada par de alunos entregue por escrito a resposta fundamentada à questão 8 da Parte I e à questão 4 da Parte II. O professor poderá fornecer um feedback escrito a cada um dos pares, podendo ainda incluir uma ou duas perguntas sobre o trabalho realizado.

Os alunos poderão realizar uma autoavaliação, procedendo a uma reflexão individual sobre as aprendizagens realizadas e estratégias de melhoria.

PROPOSTA DE RESOLUÇÃO

PARTE I

1. O estudo estatístico feito pela DECO foi uma sondagem. A população em estudo é composta pelos pneus de automóveis, de 25 marcas diferentes, utilizados ou em utilização.

2. Não temos informação suficiente para saber se a amostra é representativa. Os dados foram obtidos por inquérito a 37 mil utilizadores de pneus, que incluem consumidores portugueses, espanhóis, italianos e belgas. Não é dito qual foi o processo de amostragem, em particular, se os consumidores inquiridos nos 4 países foram selecionados aleatoriamente. Por outro lado, apesar de se tratar de uma amostra de grande dimensão, é de esperar que noutros países do mundo existam igualmente consumidores de pneus das 25 marcas estudadas. Não se sabe, por exemplo, se os condutores noruegueses ou os condutores australianos teriam uma experiência diferente de utilização de pneus que alterasse significativamente os resultados da sondagem.

3. A variável estatística é a durabilidade dos pneus de cada marca, em quilómetros feitos. Esta variável é quantitativa e contínua. (**Nota:** Os dados com que estamos a trabalhar são resultantes dos dados recolhidos e tratados pela DECO sobre a durabilidade dos pneus).

4. Foi usada a folha de cálculo (ou a calculadora gráfica) para a obtenção da média e da mediana:

$$\bar{x} = 40013$$

$$M_e = 41058$$

Os resultados mostram que a média das durações dos pneus das 24 marcas é cerca de 40 mil quilómetros. A mediana dá a informação de que 12 das marcas oferecem pneus com uma durabilidade superior ou igual a 41058 quilómetros e 12 das marcas oferecem pneus com uma durabilidade inferior ou igual a 41058 quilómetros.

4.1. Verificamos que a média é inferior à mediana ($\bar{x} < M_e$). Isto leva a pensar que a distribuição dos dados poderá ser assimétrica, com um enviesamento à esquerda.

4.2. Sendo a mediana igual a 41058, podemos afirmar que 50% das marcas oferecem pneus com uma durabilidade superior ou igual a 41058 quilómetros, logo superior a 40000 quilómetros. Portanto o critério de expectativa de vida muito longa do pneu é cumprido por, pelo menos, metade das marcas. Examinando os dados, verifica-se que há 15 marcas que cumprem o critério referido.

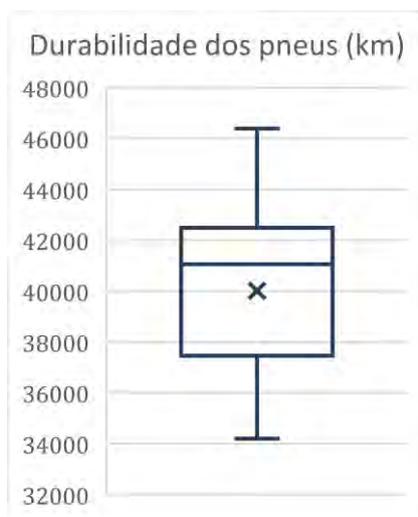
5. Foram calculadas na folha de cálculo (ou calculadora gráfica) a amplitude e a amplitude interquartil:

Mínimo	34210
Máximo	46393
Amplitude	12183
Mediana	41058
1º Quartil	37489
3º Quartil	42371
$Q_3 - Q_1$	4882

5.1. Observa-se uma amplitude de 12183 quilómetros, indicando uma considerável diferença entre a marca que tem maior durabilidade e a que tem menor durabilidade, o que significa uma grande dispersão. A diferença de durabilidade entre pneus de marcas diferentes ultrapassa efetivamente os 12 mil quilómetros, como é referido no artigo do Economista, onde se afirma: “Um estudo realizado pela Deco Proteste concluiu que a escolha dos pneus corretos pode ajudar o seu carro a percorrer mais 12 mil quilómetros do que a marca com pior avaliação”.

A amplitude interquartil é de 4882 quilómetros, pelo que se pode concluir que os 50% de dados centrais (entre o 1º quartil e o 3º quartil) estão relativamente concentrados, embora isso não signifique que estejam distribuídos de forma simétrica.

6. Foi construído no Excel o diagrama em caixa de bigodes (a média está assinalada com ×).



6.1. Como mostra o diagrama em caixa de bigodes, a média é inferior à mediana, o que significa uma assimetria com enviesamento à esquerda, mas que só se verifica na parte central dos dados, como é revelado pela representação gráfica. Como se pode verificar, há uma maior concentração de dados entre a mediana e o 3º quartil (25% dos dados concentrados num intervalo de amplitude 1313 km) do que entre o 1º quartil e a mediana (25% dos dados concentrados num intervalo de amplitude 3569 km).

6.2. A distribuição dos dados é assimétrica, com enviesamento à esquerda na parte central dos dados-

7. O valor obtido no Excel (ou calculadora) para o desvio padrão foi $s = 3316$. Esta medida é um indicador da dispersão dos dados em torno da média, que tem o valor $\bar{x} = 40013$. Para discutir a afirmação citada (*Relativamente à durabilidade média dos pneus considerados, há uma dispersão, em quilómetros, aproximadamente igual a uma viagem de Lisboa a Varsóvia*), foi feita uma consulta no site ViaMichelin para obter a distância de uma viagem de carro de Lisboa a Varsóvia. De acordo com o trajeto recomendado pelo site, serão percorridos 3324 km entre a origem e o destino, um valor próximo do desvio padrão. Portanto, a distância entre Lisboa e Varsóvia, em quilómetros, é uma boa estimativa para a variabilidade da durabilidade dos pneus das várias marcas relativamente à média.

8. Atendendo aos resultados obtidos, é possível concluir que a maioria das marcas estudadas oferece pneus que têm uma durabilidade acima da considerada desejável (40 mil quilómetros). Verificou-se, ainda, que pelo menos 50% das marcas oferece uma durabilidade superior a 41 mil quilómetros, de acordo com a mediana obtida. Há uma grande concentração dos dados entre a mediana e o 3º quartil, o que mostra que esses 25% dos dados estão bastante próximos, sendo pouco acentuadas as diferenças entre eles. Assim, se não existirem outros critérios a considerar para a recomendação a

fazer (por exemplo: o preço, a rapidez na entrega e montagem, a proximidade de um fornecedor, a confiança na marca, etc.), uma hipótese de resposta seria a seguinte:

- *maior durabilidade* (valor superior ou igual a Q_3), por ordem de preferência: 1º Michelin, 2º BF Goodrich, 3º Continental, 4º Bridgestone, 5º Goodyear, 6º Dunlop.

PARTE II

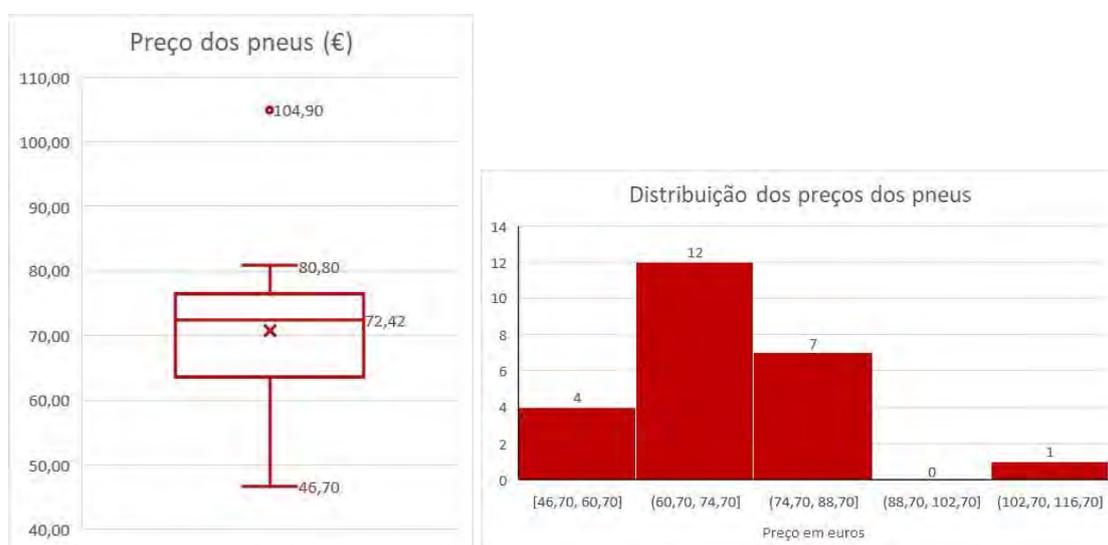
1. Foram obtidas no Excel (ou calculadora gráfica) as seguintes medidas de localização e dispersão para a distribuição dos preços dos pneus (do tipo indicado):

Mínimo	46,70
Máximo	104,90
Amplitude	58,20
Média	70,74
Mediana	72,42
1º Quartil	64,28
3º Quartil	75,91
Desvio Padrão	11,50

A média dos preços praticados pelas marcas, para um pneu do tipo considerado, é de 70,74€. Isto significa que um consumidor, caso decida optar por uma das 24 marcas de pneus, pode considerar esse valor como o preço que, em média, lhe custará 1 pneu. A mediana é superior à média e igual a 72,42€. Portanto, 50% das marcas consideradas cobram, pelo pneu, um preço superior ou igual a 72,42€ e 50% das marcas cobram um preço inferior ou igual àquele.

A amplitude é igual a 58,20€, o que significa que a diferença de preços entre marcas, para o mesmo pneu, pode ser superior a cinquenta euros. O desvio padrão indica uma dispersão de 11,50€, em relação à média.

2. Foi construído o diagrama em caixa de bigodes e também o histograma para a distribuição dos preços dos pneus:



2.1. A existência de um valor atípico é evidenciada nas duas representações (trata-se do máximo, que é igual a 104,90€).

Repare-se, ainda, que a distribuição dos dados apresenta uma assimetria esquerda, tanto na parte central, como nas “caudas”.

O histograma mostra 5 classes de amplitude igual a 14 euros, sendo que a penúltima das classes tem 0 dados e a última classe contém apenas 1 dos dados. Os restantes estão distribuídos por 3 classes. O

único dado que faz parte da classe definida pelo intervalo $]102,70; 116,70]$ é o máximo e constitui um *outlier*.

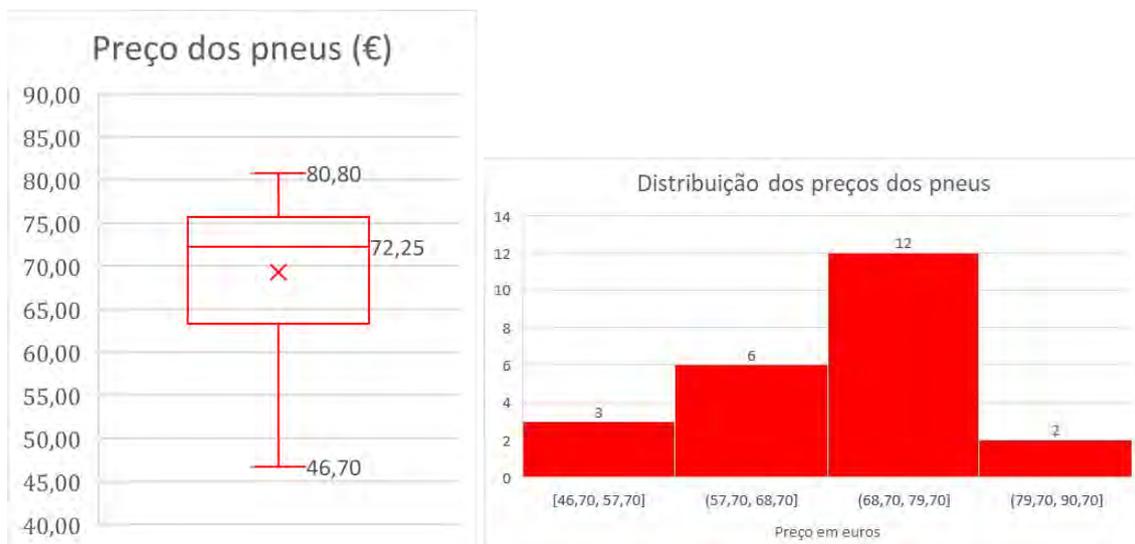
Observa-se também que os 50% dos pneus mais baratos têm preços entre 46,70€ e 72,42€, ou seja, uma amplitude de preços de cerca de 26 euros; por outro lado, os 50% dos pneus mais caros têm preços entre 72,42€ e 104,90€, ou seja, uma amplitude de preços de cerca de 32 euros.

2.2. Foi observada a existência de um *outlier* (o preço do pneu da marca Michelin). A seguir, foi retirado o *outlier* e encontrados o mínimo, o máximo e a mediana.

Mínimo	46,70
Máximo	80,80
Mediana	72,25

Há uma evidente alteração no máximo (que era igual a 104,90), mas uma pequena alteração na mediana (que era igual a 72,42), o que se explica pela resistência da mediana à alteração dos valores extremos, ao contrário da média.

Depois de retirado o *outlier*, foram obtidos o diagrama em caixa de bigodes e o histograma:



Repare-se que, como no caso anterior, mantém-se a assimetria esquerda, notando-se que a média se afasta mais da mediana, pois já não é influenciada pelo *outlier* que foi retirado, estando agora a ser influenciada pelos valores que provocam o enviesamento da distribuição dos dados.

Analisando o histograma, nota-se que os dados estão agora distribuídos em 4 classes (em vez de 5) com uma amplitude de 11 euros e que a classe modal, que contém 12 dos preços considerados, corresponde ao intervalo $]68,70; 79,70]$.

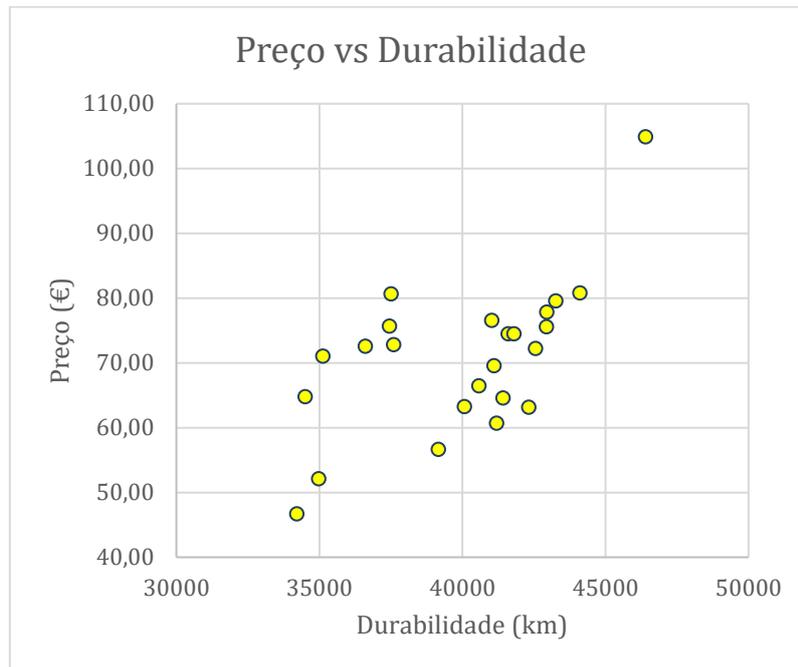
Calcularam-se, ainda, as seguintes diferenças, que são evidenciadas no diagrama em caixa de bigodes:

$$\text{Máx} - M_e = 8,55$$

$$M_e - \text{Mín} = 25,55$$

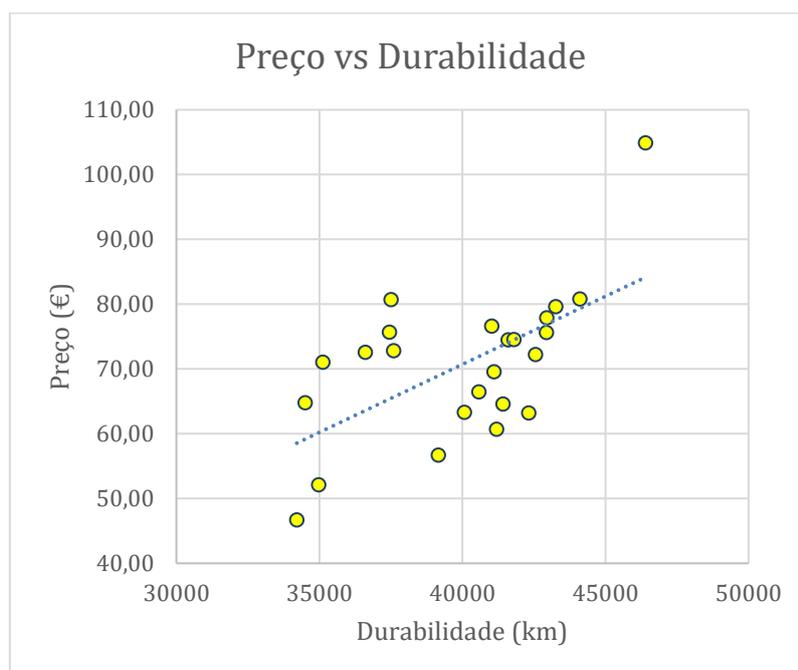
Assim, nesta situação, verifica-se que os 50% dos pneus mais baratos têm preços entre 46,70€ e 72,25€, ou seja, uma amplitude de cerca de 26 euros (não houve alteração); por outro lado, os 50% dos pneus mais caros têm preços entre 72,25€ e 80,80€, ou seja, uma amplitude de cerca de 9 euros (houve uma alteração considerável).

3. Recorrendo ao Excel, foi construído o diagrama de dispersão, com **todos os dados**. Considerou-se a variável *durabilidade* como independente e a variável *preço* como dependente.

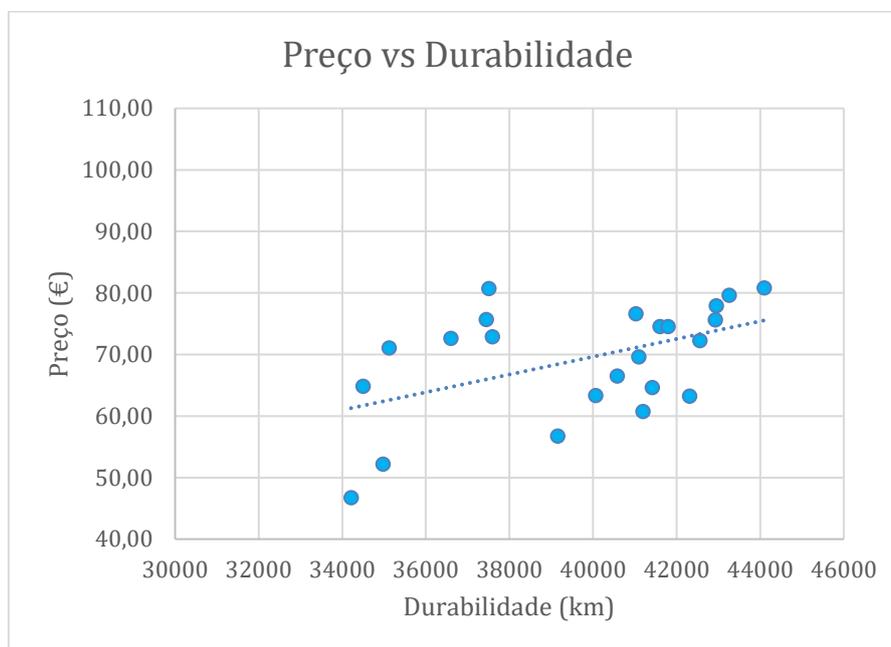


A nuvem de pontos que se observa no gráfico é indicadora de uma associação linear positiva entre as variáveis, embora, aparentemente, esta não seja forte pois os pontos da nuvem não se concentram sobre uma reta. Calculando o coeficiente de correlação no Excel (ou na calculadora gráfica), obtém-se o valor $r = 0,61$, que indica uma associação linear positiva fraca/moderada entre a durabilidade do pneu e o seu preço, quando se consideram os dados das 24 marcas.

Como esperávamos, tendo em conta o resultado anterior para o coeficiente de correlação, a reta de regressão não é um bom modelo para representar a associação entre as variáveis em estudo.



Mesmo assim, decidiu-se examinar a situação em que se retira o *outlier*, isto é, em que se excluem os dados relativos à marca Michelin. Nesse caso, o diagrama de dispersão revela ainda uma associação positiva, mas a nuvem de pontos está aparentemente mais dispersa. O coeficiente de correlação é, agora, $r = 0,49$. Assim, se a marca Michelin for ignorada, observa-se uma associação linear positiva, ainda mais fraca, entre o preço e a durabilidade do pneu.



4. A análise efetuada permite concluir que a marca Michelin é “um caso à parte”. Significa um *outlier* para a variável preço, embora não o seja para a variável durabilidade.

Como se disse anteriormente, verifica-se que a correlação entre a durabilidade e o preço é fraca, pelo que não podemos utilizar este critério para a escolha do “melhor” pneu. Tomaremos um outro critério, expresso pela razão entre a durabilidade e o preço (que dá o número de quilómetros percorridos por cada euro do valor de um pneu). Segue-se a tabela com os valores dessa razão, considerando apenas as 5 marcas que anteriormente (na Parte I) foram seleccionadas (com base na durabilidade), depois de retirada a marca Michelin:

	Durabilidade (km)	Preço (€)	Durab./Preço (km/€)
Dunlop	42551	72,25	588,94
Goodyear	42927	75,61	567,74
Bridgestone	42944	77,88	551,41
BF Goodrich	44096	80,80	545,74
Continental	43258	79,60	543,44

A partir da tabela, e assumindo que o mais vantajoso será o maior valor da razão entre a durabilidade e o preço, uma possível recomendação seria, então, a seguinte:

- *melhor razão durabilidade-preço* (por ordem de preferência): 1º Dunlop, 2º Goodyear, 3º Bridgestone, 4º BF Goodrich, 5º Continental.

É importante notar que esta recomendação tem por base uma recolha de preços de pneus realizada apenas num site da Internet, pelo que outras pesquisas poderiam revelar preços diferentes que modificariam as conclusões obtidas.

ÍNDICE BIG MAC

TRABALHO DE PROJETO

Big Mac
??? €



O Índice Big Mac, oficialmente Big Mac Index, é um índice calculado sobre o preço do Big Mac, da cadeia de fast food McDonald's, em mais de cem países. O índice foi criado, em 1986, pelo jornal *The Economist* para explicar um conceito económico chamado Paridade de Poder de Compra.

Podes até nem gostar de hambúrgueres, mas seguramente já entraste num McDonald's. A cadeia McDonald's está presente em mais de 120 países, com mais de 36.000 restaurantes, tendo deste modo uma grande importância.

Fã ou não deste tipo de alimentação 'fast food', é interessante perceber qual é o preço do Big Mac – um dos produtos mais antigos e consumidos nesta cadeia – nos diferentes países do mundo onde a McDonald's está presente e o seu significado.

Propomos que desenvolvas um Trabalho de Projeto baseado neste índice, recorrendo aos teus conhecimentos de Estatística e/ou de outras áreas que consideres relevantes.

NOTAS PARA O PROFESSOR

Podemos definir o trabalho de projeto como uma metodologia de aprendizagem que envolve os alunos na aprendizagem de conhecimentos e competências através de um processo de investigação alargado, estruturado em torno de questões complexas, autênticas e de produtos e tarefas cuidadosamente concebidos.

O trabalho de projeto permite ainda que os alunos aprendam na escola e, em particular, na disciplina de matemática sobre temas relevantes para a sua vida e para a sociedade ao mesmo tempo que desenvolvem diversas competências. Deste modo, podemos afirmar que através do desenvolvimento de um projeto os alunos adquiram os objetivos de aprendizagem e as competências do Perfil dos Alunos à Saída da Escolaridade Obrigatória.

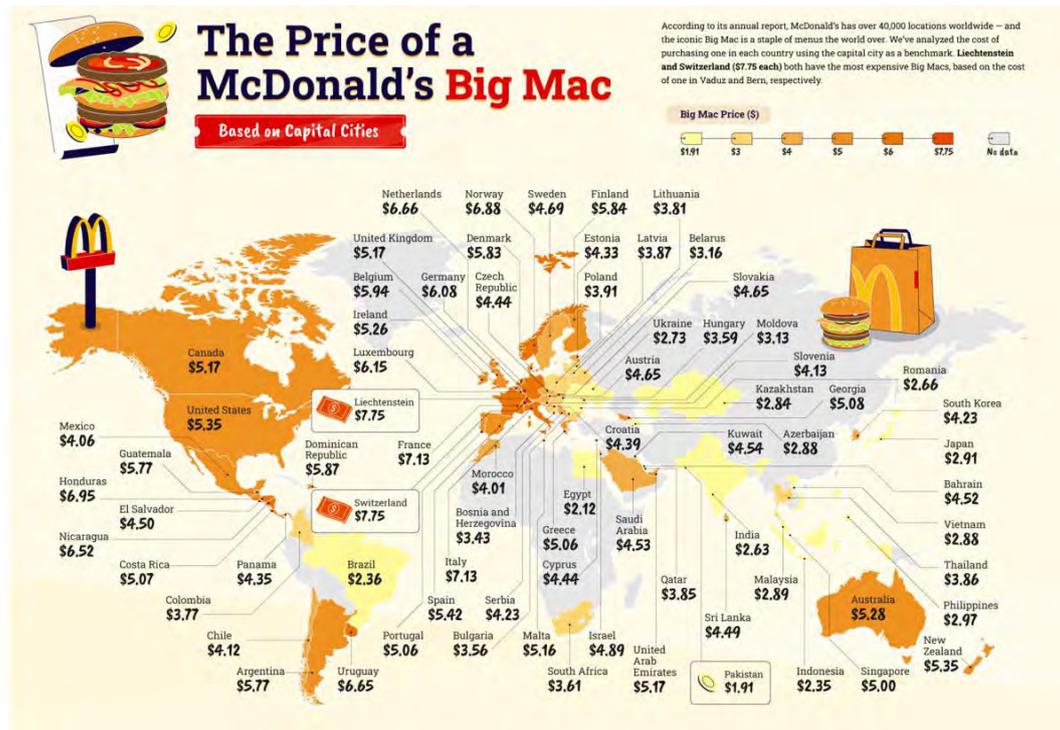
Uma questão prévia, que deve ser pensada pelo professor, é a duração do projeto, tanto no número de aulas que pode disponibilizar para as várias etapas como o período temporal de realização do projeto.

O desenvolvimento de um Projeto pressupõe as seguintes etapas que podem ocorrer dentro e fora da sala de aula:

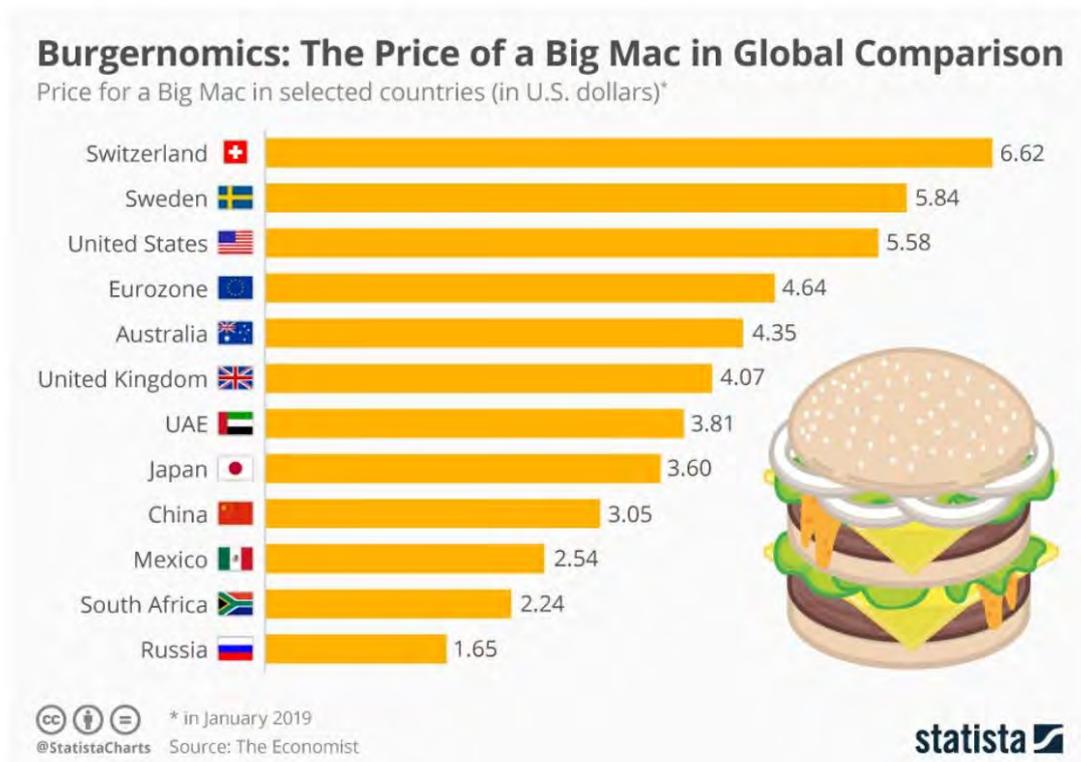
1ª Etapa: Formular uma questão ou um problema a estudar

Nesta etapa é necessário discutir com os alunos acerca do que é o Índice Big Mac. Este momento requer um trabalho de pesquisa e será importante registar as ideias que os alunos vão sugerindo para encontrar a questão ou o problema a estudar. Essa questão pode ser muito diferente de turma para turma ou de grupo para grupo.

Como resultado desse trabalho de pesquisa, poderão surgir descrições, dados, imagens, gráficos, etc., como se ilustra nos exemplos seguintes de representações visuais:



Fonte: https://pt.wikipedia.org/wiki/Ficheiro:01_The-Price-of-McDonalds-Big-Mac_World-Map.png



Fonte: <https://www.idealista.pt/news/financas/economia/2019/01/22/38540-visita-guiada-ao-mundo-dos-big-mac>

4ª Etapa: Executar

Trata-se da etapa destinada a analisar os dados recolhidos com vista a responder ao problema inicialmente formulado. A análise estatística deve centrar-se nos processos e nas representações que sejam consideradas adequadas para dar resposta ao problema ou questões.

5ª Etapa: Monitorizar

Ao longo do desenvolvimento do projeto deve existir uma monitorização sistemática dos progressos do trabalho, pelos alunos e pelo professor. Os alunos devem ser incentivados a refletir e analisar os vários passos e verificar se há necessidade de fazer ajustamentos e avaliar a forma de o apresentar. Esta etapa pode ocorrer várias vezes durante o projeto, pelo que o professor deve reservar alguns momentos da aula para este efeito. O feedback dado pelo professor durante este processo pode ser decisivo para o sucesso do projeto.

6ª Etapa: Apresentar

É a concretização de uma etapa planeada. Os alunos devem elaborar a sua apresentação (PowerPoint, Prezi, vídeo, cartaz, poster, folheto, brochura, ...) de acordo com o que foi decidido anteriormente no plano de trabalho. Deve ser solicitado aos alunos a elaboração de um pequeno relatório escrito.

7ª Etapa: Avaliar

A avaliação é parte integrante do trabalho de projeto, deste modo a avaliação de todo o trabalho realizado deve ser incluída nas conclusões ou nas recomendações finais. Nesta avaliação os alunos devem ter em consideração os objetivos iniciais e fazer um balanço da forma como foram atingidos, indicar a existência de limitações ou dificuldades e apresentar propostas de possíveis investigações futuras.

É recomendável que os alunos respondam a um questionário de autoavaliação das aprendizagens e das competências desenvolvidas.

DIAS QUENTES

TRABALHO DE PROJETO



O Serviço Copernicus para as Alterações Climáticas (C3S) apoia a sociedade, fornecendo informações fidedignas sobre o clima passado, presente e futuro na Europa e no resto do mundo. No site deste organismo europeu foram publicadas algumas conclusões breves sobre o ano de 2022:

- Para uma grande parte da Europa, o ano de 2022 foi o mais quente de que há registo.
- Os extremos de calor no final da primavera e no verão resultaram em condições perigosas para a saúde humana.
- O Sul da Europa registou um número recorde de dias com «stress térmico muito forte» durante o verão, consistente com uma tendência crescente no número de dias com «stress térmico forte» ou «muito forte» em toda a Europa e uma tendência decrescente no número de dias «sem stress térmico».

Conheces o significado científico da expressão “anomalia climática”?

No site da Agência Europeia do Ambiente, pode ler-se a seguinte informação³⁵:

Os dias quentes são definidos como estando acima do percentil 90 da temperatura máxima diária (relativamente a um dado período de referência).

Propomos que desenvolvas um Trabalho de Projeto sobre a ocorrência de anomalias climáticas, recorrendo aos teus conhecimentos de Estatística e de outras áreas que consideres relevantes e necessárias. Poderás considerar a variável temperatura ou a variável precipitação. Uma possibilidade será estudar dados climatológicos de Portugal e outra será fazer comparações entre diferentes países, por exemplo, da Europa.

³⁵ <https://www.eea.europa.eu/data-and-maps/figures/changes-in-duration-of-warm-spells-in-summer-across-europe-in-the-period-1976-2006-in-days-per-decade-10>

NOTAS PARA O PROFESSOR

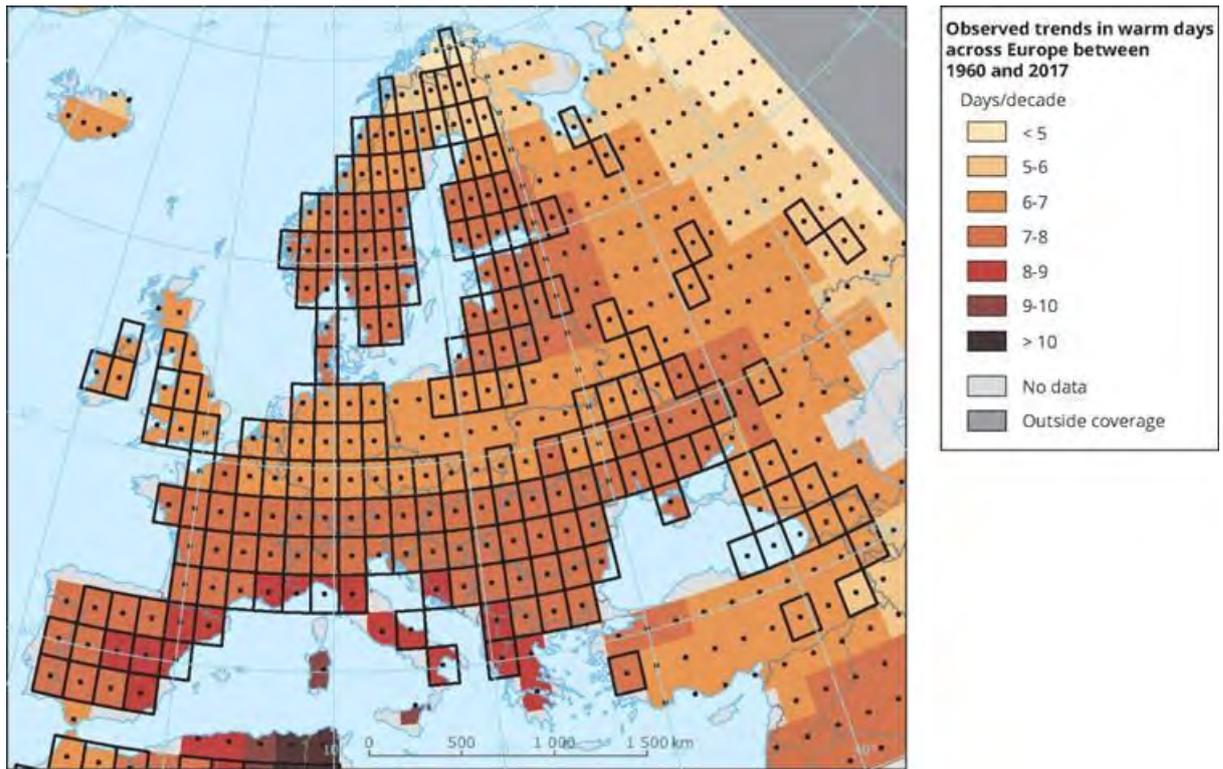
O trabalho de Projeto é uma metodologia de aprendizagem fundamental na educação matemática dos jovens do século XXI. Esta metodologia está em consonância com o que defendido no Perfil dos Alunos à Saída da Escolaridade Obrigatória que tem como pressuposto de que cada área curricular deve contribuir para o desenvolvimento de todas as áreas de competências consideradas. Deste modo, a implementação de um Trabalho de Projeto na disciplina de matemática reveste-se enorme importância para que os alunos mobilizem e desenvolvam valores e competências que lhes permitem intervir na vida e na sociedade, tomando decisões fundamentadas sobre questões diversas, nomeadamente sobre as alterações climáticas, e tenham a capacidade de participar de forma ativa, consciente e responsável na sociedade.

O trabalho de Projeto tem sempre um carácter desafiante e inovador e inovador e permite que os alunos tenham oportunidade de escolher uma ideia sobre a qual pretendem trabalhar. Assim sendo, não é possível apresentar uma proposta de resolução, mas apenas algumas sugestões e a noção de que o desenvolvimento de um Projeto pressupõe as seguintes etapas:

1ª Etapa: Formular uma questão ou um problema a estudar

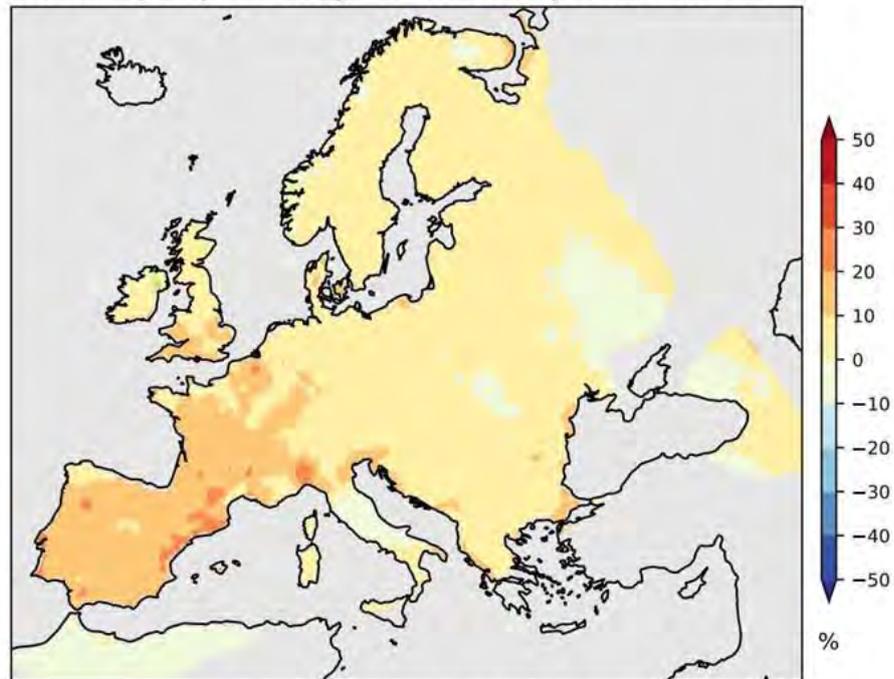
Nesta etapa é necessário discutir com os alunos o tema das alterações climáticas e do aquecimento global que têm provocado, ano após ano, temperaturas mais elevadas. Compreender o que são anomalias climáticas, como se definem, e as suas implicações sociais, designadamente na saúde humana, sobre o ambiente e sobre as atividades económicas é fundamental para a formular o problema e a questão de investigação. Posto isto, é necessário selecionar a(s) variável(eis) climatológica(s) a estudar para abordar a questão formulada e responder ao problema. Este é um tema muito abrangente, razão pela qual é fundamental uma discussão prévia para que os alunos clarifiquem as suas ideias e formulem uma questão ou problema de forma tão objetiva quanto possível. Poderá ser necessário envolver outras áreas nesta discussão. O professor deve antecipadamente analisar esta situação em função do tempo que irá dedicar ao desenvolvimento do projeto.

Uma possível abordagem inicial poderá consistir na apresentação de algumas imagens que contribuam para despertar o interesse e a discussão em torno do tema. Apresentam-se algumas imagens e respetivos sites da Internet que poderão servir esse propósito.



Fonte: <https://www.eea.europa.eu/data-and-maps/figures/changes-in-duration-of-warm-spells-in-summer-across-europe-in-the-period-1976-2006-in-days-per-decade-10>

Anomaly in percentage of 'warm day-times' in 2022

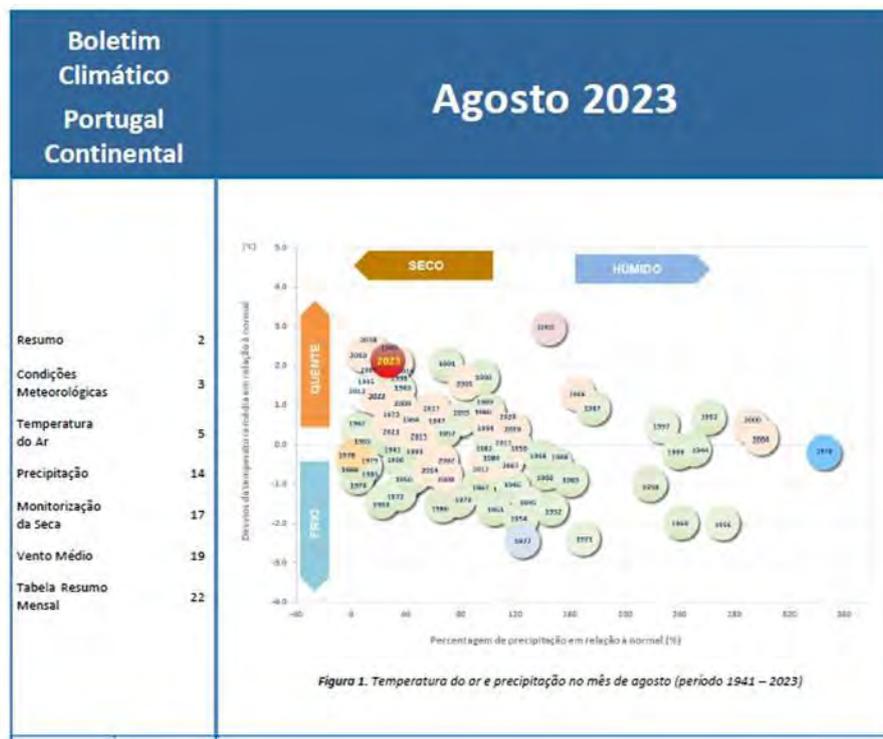


Data source: E-OBS Credit: C3S/KNMI Reference Period: 1991-2020

Fonte: <https://climate.copernicus.eu/esotc/2022/extreme-heat>



Fonte: <https://www.ipma.pt/pt/oclima/monitoriza.dia/>



Fonte:

https://www.ipma.pt/resources.www/docs/im.publicacoes/edicoes.online/20230908/FFXYNHtFtktOuzKSGRjD/cli_20230801_20230831_pcl_mm_co_pt.pdf

2ª Etapa: Estabelecer um plano de trabalho

Após a formulação do problema, e tendo em conta o tempo disponível para desenvolver o projeto, é necessário estabelecer um plano de trabalho. Este plano deve prever de forma detalhada o processo de recolha de dados, assim como os diversos recursos necessários e as atividades para a realização do projeto com sucesso. Nesta etapa deve considerar a forma de apresentar e/ou divulgar os resultados do projeto. Assim, tomadas estas decisões, é possível planear o estudo estatístico a realizar que pode envolver os conhecimentos estatísticos dos alunos, mas pode envolver a aprendizagem de novos conhecimentos.

3ª Etapa: Recolher informação

A recolha de informação é uma etapa decisiva num projeto. Nesta etapa os alunos já definiram as suas questões e as variáveis em estudo. Para a recolha de dados poderão recorrer aos sites oficiais de vários organismos nacionais ou internacionais onde existem dados que lhes permitam estudar o problema em estudo. Outra possibilidade é os próprios alunos se envolverem na recolha de dados nas suas cidades ou regiões, tal ficará sempre dependente do problema em estudo que devido à amplitude do tema pode assumir muitas possibilidades.

Nota: Vários conjuntos de dados fiáveis são disponibilizados pelo Instituto Português do Mar e da Atmosfera (IPMA). Também poderão ser encontrados dados úteis em organismos internacionais, como a Agência Europeia do Ambiente ou o Serviço Copernicus (C3S).

4ª Etapa: Executar

A execução do projeto implica fazer a análise da informação e dos dados recolhidos com vista a responder ao problema inicialmente formulado. A análise estatística deverá incluir os processos e as representações mais adequadas para dar resposta ao problema inicialmente formulado. Os alunos devem recorrer a diferentes formas de representar os seus dados, escolhendo e justificando as que consideram mais adequadas à situação.

5ª Etapa: Monitorizar

O processo de monitorização do projeto deve ocorrer ao longo do desenvolvimento do projeto. O professor desempenha um papel muito importante neste processo, colocando questões que permitam aos alunos rever sistematicamente os progressos do trabalho, verificando se há necessidade de fazer ajustamentos ao plano estabelecido e tendo em conta a forma de apresentar o projeto. Sempre que necessário o professor deve oferecer feedback aos alunos que contribua para promover as aprendizagens.

6ª Etapa: Apresentar

A forma de apresentar o projeto foi pensada na segunda etapa e ao longo das etapas seguintes os alunos trabalharão tendo presente esta etapa. A apresentação pode tomar diferentes formas, nomeadamente em formato digital, recorrendo ao PowerPoint, Prezi, vídeo ou outro, ou através de um cartaz, poster, folheto, brochura de acordo com o que foi decidido no plano de trabalho. É muito importante que seja produzido um relatório escrito deste trabalho.

7ª Etapa: Avaliar

A avaliação é parte integrante do trabalho de projeto. Deste modo, deve estar presente em todas as etapas do trabalho, devendo igualmente ser incluída nas conclusões ou nas recomendações finais. É muito importante que os alunos reconheçam e reflitam sobre as diversas aprendizagens realizadas quer seja de conhecimentos quer das competências desenvolvidas. Desta forma, é necessário levar os alunos a refletirem sobre o trabalho e a forma como este foi desenvolvido. É recomendável apresentar aos alunos um questionário de autoavaliação onde eles tenham oportunidade de reconhecer as aprendizagens e as competências desenvolvidas assim como reconhecer as vantagens de um trabalho desta natureza na sua formação pessoal e académica. Nesta avaliação é importante ter como referência os objetivos iniciais que foram estabelecidos e perceber como foram atingidos. É ainda relevante que os alunos apontem dificuldades sentidas e a existência de limitações; por fim, os alunos poderão apresentar propostas de possíveis investigações futuras.

A concluir

Esta publicação pretende servir de apoio ao professor na leção do tema de Estatística, comum a todas as disciplinas de matemática do Ensino Secundário: Matemática A, Matemática B, Matemática Aplicada às Ciências Sociais e Matemática dos Cursos Profissionais.

O conjunto de tarefas aqui disponibilizado abrange todos os tópicos e subtópicos do programa e cada tarefa, para além de uma proposta de resolução, tem notas metodológicas para apoio ao professor, onde se encontra o número de aulas de 50 minutos recomendado para cada tarefa, que sistematizamos na tabela seguinte:

Duração prevista para a realização da tarefa (em tempos de 50 minutos)

DESIGNAÇÃO DA TAREFA	N.º DE AULAS (50 MINUTOS)
"I have a dream"	2
A evolução na vida da mulher portuguesa nos últimos 50 anos	2
Doçura ou travessura (e a questão da desigualdade)	2
Consumo de água	4
Quanto estica um elástico?	3
Compreender um estudo sobre hipertensão nos jovens	2
O coeficiente de correlação linear	3
Quanto mais poluição, melhor?	2
Na hora de substituir os pneus do carro	3
Trabalho de projeto – Índice Big Mac	≥ 5
Trabalho de projeto – Dias quentes	
TOTAL	≥ 28

Os autores desta *brochura* partilham a visão do projeto ALEA (Ação Local de Estatística Aplicada), disponível em www.alea.pt, a completar em breve 25 anos no cumprimento da sua missão. Neste portal web, encontram-se materiais didáticos de apoio à leção e de promoção da Literacia Estatística, entre os quais os dossiês didáticos, ActivAlea's e TarefAlea's. Recomendamos, em particular, a consulta das seguintes TarefAlea's, que complementam as tarefas propostas nesta *brochura* e estão disponíveis para download:

- O envelhecimento da população da União Europeia (EU-27), com destaque para a população portuguesa (disponível [AQUI](#))
- A desigualdade salarial entre homens e mulheres (disponível [AQUI](#))

A primeira destas tarefas envolve o tópico *Dados Univariados* e os subtópicos *Medidas de Localização* e *Medidas de Dispersão* e o tópico *Dados Bivariados*, no subtópico *Gráfico de Linhas*. A segunda tarefa aborda o tópico *Dados Bivariados* e os subtópicos *Diagrama de Dispersão*, *Coefficiente de Correlação Linear* e *Reta de Regressão*. Privilegia-se, ainda, nas duas tarefas, a interpretação de diversos tipos de gráficos e o cuidado a ter nessa interpretação, já que alguns gráficos podem transmitir uma ideia errada sobre os dados.

Reafirmamos a importância da utilização sistemática da tecnologia no ensino e na aprendizagem da Estatística, como está explícito nas Aprendizagens Essenciais de Matemática para o ensino secundário (cf., em particular, as ideias-chave das Aprendizagens Essenciais). Nesta *brochura* privilegiámos o recurso à folha de cálculo Excel, mas consideramos igualmente oportuno o uso da calculadora gráfica, podendo ambos os recursos coexistir na sala de aula. Uma outra possibilidade é a utilização do GeoGebra, com o qual podem ser construídas representações dinâmicas e realizadas análises estatísticas, havendo aliás vários recursos acessíveis na Internet, produzidos pela comunidade GeoGebra. Outros aplicativos disponíveis online são igualmente adequados para diversas explorações e ilustrações de conceitos estatísticos, entre os quais, sugerimos os aplicativos da coleção Rossman/Chance (<https://www.rossmanchance.com/applets/index2021.html>) e os aplicativos disponíveis no website StatCrunch (<https://www.statcrunch.com/>).

Salientamos, ainda, o valor e a relevância da utilização de dados autênticos e reais nas explorações estatísticas e no desenvolvimento da literacia estatística, contribuindo para uma cidadania crítica e informada. Existem atualmente muitos dados acessíveis na Internet que são disponibilizados por diversos organismos, nacionais e internacionais. Estes permitem ao professor a construção de tarefas com significado económico e social e dão a oportunidade de dinamizar trabalhos de projeto em temas pertinentes e atuais escolhidos pelos alunos. Na presente *brochura*, foram usados, principalmente, dados provenientes do PORDATA (Estatísticas sobre Portugal e a Europa), mas diversos sites como o INE (Instituto Nacional de Estatística), o IPMA (Instituto Português do Mar e da Atmosfera), o BPstat (Banco de Portugal), bem como os websites *Our World in Data* (<https://ourworldindata.org/>) e *Gapminder* (<https://www.gapminder.org/>) também disponibilizam dados que podem ser usados na aula de Matemática, permitindo a interdisciplinaridade com outras áreas do saber.

A concluir, desejamos aos colegas e aos seus estudantes um excelente trabalho, em prol do sucesso das aprendizagens e do gosto e interesse pela Estatística.

Créditos das imagens

Nesta *brochura* foram utilizadas imagens de acesso gratuito, seguindo-se a respetiva atribuição de créditos:

Capa e ficha técnica:

<https://www.flaticon.com/free-icons/box-plot>

Pág. 6:

https://commons.wikimedia.org/wiki/File:Martin_Luther_King_Jr._addresses_a_crowd_from_the_steps_of_the_Lincoln_Memorial,_USMC-09611.jpg#file

Pág. 24:

<https://pt.vecteezy.com/vetor-gratis/personagem>

Pág. 40:

https://br.freepik.com/vetores-gratis/saco-de-doces-de-halloween-de-design-plano_9700692.htm

Pág. 54:

<https://portaldaagua.pt/materiais-de-comunicacao/agua-e-vida/>

Pág. 68:

<https://unsplash.com/pt-br/fotografias/pessoa-segurando-borracha-t-HffGSnyE>

Pág. 79:

https://br.freepik.com/vetores-gratis/medico-medindo-a-pressao-arterial-para-paciente-do-sexo-masculino-medica-sentada-a-mesa-na-clinica-ou-hospital-e-verificando-a-pressao-arterial-da-ilustracao-vetorial-plana-do-homem-doente-cardiologia-conceito-de-saude_26921680.htm

Pág. 93:

<https://depositphotos.com/photos/correlation.html?qview=48041027>

Pág. 114:

https://www.freepik.com/free-vector/scene-with-factory-buildings-tow-truck-site_6822419.htm

Pág. 121:

<https://pixabay.com/pt/photos/pneus-c%C3%ADrculos-revendedor-de-pneus-2989872/>

Pág. 136:

https://pt.wikipedia.org/wiki/%C3%8Dndice_Big_Mac#/media/Ficheiro:Big_Mac_hamburger.jpg

Pág. 141:

https://br.freepik.com/vetores-gratis/ilustracao-de-calor-de-verao-plana-com-homem-na-frente-do-ventilador-e-ar-condicionado_27995492.htm

ÍNDICE REMISSIVO

A

amostra
aleatória, 7, 14, 16, 18, 19, 20, 23
aleatória simples, 7, 14, 16, 19
dimensão da, 7, 9, 14, 15, 16, 17, 19, 20, 21, 22,
23, 71, 129
enviesada, 13, 14, 19
não aleatória, 19
por conveniência, 14
representativa, 14, 19, 122, 129
amostragem, 14, 19, 129
amplitude, 7, 13, 43, 57, 58, 59, 78, 86, 123, 127,
129, 130
amplitude interquartil, 13, 43, 78, 123, 126, 127,
129, 130
Anscombe, 101
associação linear, 93, 100, 105, 106, 107, 109, 111,
112, 113, 117, 119, 120, 125, 134, 135

C

calculadora gráfica, 3, 15, 44, 58, 59, 60, 70, 124,
127, 128, 129, 132, 134, 148
censo, 28, 37
classe, 36, 76, 132, 133
amplitude de, 64, 126, 132, 133
modal, 133
coeficiente de correlação (linear), 79, 87, 93, 94,
95, 97, 98, 99, 100, 101, 105, 106, 107, 109,
112, 113, 115, 117, 118, 119, 120, 125, 127,
134, 135, 147, 148
correlação, 79, 87, 93, 94, 95, 97, 98, 99, 100, 101,
103, 105, 106, 107, 108, 109, 110, 111, 112,
113, 115, 117, 118, 119, 120, 125, 126, 127,
134, 135, 147
espúria, 117, 118, 120

D

dados
bivariados, 33, 58, 66, 87, 93, 94, 96, 101, 106,
111, 117, 126, 127
experimentais, 69, 71
qualitativos, 39, 87
quantitativos, 58, 126
univariados, 4, 13, 33, 43, 58, 64, 70, 86, 148
decil, 42, 51
desvio padrão, 43, 77, 78, 86, 89, 123, 126, 127,
130, 132
diagrama
de caule-e-folhas, 41, 43, 44, 47, 57, 58, 59, 64

de dispersão, 34, 58, 93, 100, 101, 105, 106,
107, 111, 112, 115, 117, 118, 119, 125, 127,
134, 135
de extremos e quartis, 21
de pontos, 8, 9, 14, 21
em caixa de bigodes, 8, 20, 21, 69, 71, 76, 123,
124, 130, 132, 133
distribuição
simétrica, 22, 23, 41, 64, 123, 130
assimétrica, 41, 48, 51, 77, 78, 129, 130, 132,
133

E

enviesamento, 19, 21, 48, 52, 64, 75, 76, 129, 130,
133
estimativa, 7, 9, 19, 20, 22, 23, 130
estudo estatístico, 4, 13, 71, 73, 79, 86, 122, 126,
129, 139, 145
Excel, 7, 8, 14, 16, 17, 18, 19, 20, 22, 23, 46, 49, 52,
54, 61, 65, 75, 76, 119, 130, 132, 134, 148

F

folha de cálculo, 3, 4, 13, 14, 15, 34, 43, 44, 57, 58,
59, 60, 64, 87, 107, 111, 119, 123, 124, 127,
128, 129, 148
frequências
absolutas, 37, 39
relativas, 37, 39, 90

G

GeoGebra, 3, 14, 94, 95, 96, 107, 112, 113, 118,
148
gráfico
circular, 52, 90
com barras, 39, 41, 51
de barras, 39
de colunas, 90
de colunas empilhadas, 90
de dispersão (ver diagrama de dispersão)
de hierarquia, 49
de linhas, 33, 34, 38, 39, 57, 58, 59, 66
de percentis, 89, 90
Treemap, 49, 50

H

histograma, 33, 75, 76, 124, 132, 133

M

máximo, 7, 63, 64, 69, 73, 74, 75, 77, 121, 125, 132,
133

média, 7, 8, 9, 13, 19, 20, 21, 23, 25, 29, 30, 38, 39,
40, 41, 43, 44, 46, 47, 48, 51, 52, 55, 57, 58, 59,
64, 65, 66, 71, 73, 75, 77, 78, 86, 89, 94, 96, 100,
103, 105, 109, 112, 122, 123, 126, 129, 130,
132, 133
mediana, 7, 13, 21, 41, 43, 44, 47, 48, 75, 77, 122,
125, 126, 129, 130, 132, 133
medidas
de dispersão, 13, 15, 33, 43, 58, 69, 75, 86, 89,
124, 126, 127, 130, 132, 148
de localização, 13, 15, 43, 58, 69, 75, 86, 89, 124,
126, 127, 132
propriedades das, 44, 86
resistentes, 77, 100, 109
mínimo, 7, 63, 64, 125, 133
modelo
de regressão, 93
de regressão linear, 115, 116

N

nuvem de pontos, 93, 106, 109, 115, 117, 119, 125,
127, 134, 135

O

outlier, 14, 21, 46, 47, 48, 51, 64, 71, 76, 109, 111,
125, 127, 133, 135

P

percentil, 13, 43, 80, 86, 87, 89, 90, 141
pictograma, 41, 46
população, 7, 9, 13, 14, 15, 16, 17, 19, 20, 21, 23,
69, 70, 73, 77, 79, 86, 89, 92, 106, 117, 122, 126,
129
problema estatístico, 13, 79, 89

Q

quartil, 13, 21, 41, 43, 48, 65, 123, 126, 130

R

relação de causa-efeito, 114, 116, 117, 118, 120
reta de regressão, 95, 97, 106, 108, 116, 117, 120,
125, 134
reta dos mínimos quadrados, 106, 115, 117

S

sondagem, 129

T

tabela de frequências, 39

U

unidade de observação, 9, 55, 61

V

valor médio, 7, 9, 19, 20, 21, 23
variabilidade, 4, 8, 9, 13, 14, 21, 33, 46, 70, 73, 77,
78, 130
variável
dependente, 106, 117, 134
explanatória (ou independente), 125
independente, 106, 117, 134
perturbadora, 120
qualitativa, 79, 80, 81, 86, 87, 90
quantitativa, 23, 61, 79, 87, 93, 106, 117, 129
resposta (ou dependente), 125